Universidade de Aveiro Departamento de Química

2011

Ana Alexandra Barrinha Cordeiro

Controlo da polidispersividade em reacções de polimerização

Polydispersion control in enzymatic polymerization of oligonucleotides

Universidade de Aveiro Departamento de Química

2011

Ana Alexandra Barrinha Cordeiro

Controlo da polidispersividade em reacções de polimerização

Polydispersion control in enzymatic polymerization of oligonucleotides

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biotecnologia, realizada sob a orientação científica do Doutor João Carlos Lima, Professor Auxiliar do Departamento de Química da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa e do Doutor Jorge Saraiva, Investigador Auxiliar do Departamento de Química da Universidade de Aveiro

Dedico este trabalho à minha família e namorado pela amizade, amor e incansável apoio que demonstraram.

o júri

presidente	Prof. Doutora Ivonne Delgadillo Giraldo professora associada com agregação da Universidade de Aveiro
orientador externo	Prof. Doutor João Carlos Lima professor auxiliar da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
orientador interno	Prof. Doutor Jorge Saraiva investigador auxiliar da Universidade de Aveiro
arguente	Prof. Doutor António Jorge Parola professor auxiliar da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Agradecimentos

Em primeiro lugar ao meu orientador, Professor Doutor João Carlos Lima, agradeço por todo o conhecimento e disponibilidade prestada para responder a qualquer dúvida relacionada com este trabalho e também por me ajudar a ultrapassar os obstáculos inerentes a qualquer projecto de investigação.

Ao Professor Doutor Jorge Saraiva, orientador interno deste trabalho, agradeço pela aceitação deste cargo, disponibilidade demonstrada e entusiasmo relativamente a este projecto.

Ao Professor Doutor Pedro Baptista agradeço pela oportunidade de trabalhar no laboratório 315, por todos os ensinamentos transmitidos e por todas as reuniões e *lab meetings* que me ajudaram a evoluir como investigadora e a progredir neste projecto.

Ao Professor Doutor Rui Oliveira agradeço pelo criação do Modelo Matemático baseado em Cinética Estocástica usado neste projecto para prever a evolução da reacção enzimática catalisada pela TdT.

A todos os meus colegas e amigos do 315: André, Gonçalo, Letícia, João Rosa, Miguel, João Conde, Bruno, Maria, Carina, Madalena e Margarida por todo o interesse e apoio que demonstraram, por acreditarem no meu trabalho, pelas inúmeras discussões científicas proporcionadas e por tornarem o trabalho árduo divertido, um muito obrigado. Muito obrigado ainda ao Artur Moro por todos os ensinamentos transmitidos sobre HPLC, irradiações e por me ambientar ao mundo da Fotoquímica.

Ao Miguel, como melhor amigo e namorado, pelo carinho e amor demonstrado, por nunca ter duvidado de mim nem do meu trabalho e por me apoiar todos os dias mesmo estando longe, muitíssimo obrigado.

Por fim um muito obrigado aos meus pais e irmão pela compreensão e carinho demostrados ao longo destes dois anos e por me ajudarem a ultrapassar mais uma etapa da minha vida.

palavras-chave

Terminal Deoxynucleotidyl Transferase, reacção de polimerização, cumarinas, libertação de nucleotídeos pela luz, cinética estocástica, *Chemical Master Equation*.

Resumo

O principal objectivo deste trabalho é o desenvolvimento de um modelo de cinética estocástica para prever a polidispersividade existente em reacções de polimerização enzimática. O facto de a polimerização ocorrer em paralelo em vários locais enzimáticos que apresentam diferentes processividades resulta na obtenção de uma distribuição de produtos com diversos tamanhos. Para melhor controlar esta polidispersividade foram tomadas as seguintes considerações: a enzima seleccionada para este estudo foi a *Terminal Deoxynucleotidyl Transferase* (TdT) que catalisa a síntese de DNA na ausência de um molde; e foram utilizados derivados de 4-metilcumarinas como *cages* de nucleotídeos tornando-os inacessíveis à enzima. Após exposição a radiação monocromática, ocorre uma quebra de ligações fosfodiéster e o nucleotídeo é libertado. O referido sistema traria a possibilidade única de síntese de oligonucleotídeos em solução com sequência definida.

Foram testadas diferentes estequiometrias de reacção, através da incorporação de desoxirribonucleotídeos e ribonucleotídeos na extremidade de um oligonucleotídeo iniciador, sendo identificados padrões de incorporação distintos para cada caso. Estudos de irradiação e controlo da reacção de incorporação com DEACM-ATP, molécula utilizada para bloquear ATP, foram efectuados com sucesso. Foi ainda estabelecida uma comparação entre os dados obtidos por modulação matemática, variando diversos parâmetros, e os vários resultados de incorporação obtidos experimentalmente onde a polidispersividade é observada. Desta, resultou a observação de diferentes velocidades de incorporação para cada nucleotídeo. Diversas hipóteses baseadas na estabilidade do complexo enzima-substrato foram propostas para justificar a observação de que o padrão de incorporação do nucleotídeo de adenina seja tão distinto dos restantes. Este facto pode estar relacionado com a função biológica da TdT devido à presença de um “lariat-like” loop que se pensa estar relacionado com o posicionamento de cada nucleotídeo.

Keywords

Terminal Deoxynucleotidyl Transferase, polimerization reaction, coumarines, light-induced decaging, stochastic kinetics, Chemical Master Equation.

Abstract

The main goal of this work is the development of a stochastic-based kinetic model to predict the polydispertion in enzymatic polymerization reactions. The fact that the polymerization occurs simultaneously in several enzymatic sites which possess different processivities results in a wide size distribution with different product sizes. To besto control the observed polydispersion, the following was considered: the enzyme selected to this study was Terminal Deoxynucleotidyl Transferase (TdT) as it does not require a template sequence to catalyze DNA synthesis; and 4-methylcoumarin derivatives were the selected molecules to act as caging agents for different nucleotides making them inaccessible to the enzyme. The nucleotide is then released by light due to the break of the phosphodiester bond between the two molecules when submitted to monochromatic radiation. The referred system would bring a unique possibility of creating oligonucleotides in solution with a desired sequence.
Different reaction stoichiometries were tested by the incorporation of deoxynucelotides and ribonucleotides into an oligonucleotide initiator, where distinct incorporation patterns were identified. Irradiation studies and controls for the incorporation reaction of DEACM-ATP, the cage molecule for ATP, were successfully performed. A comparison was established between the mathematic model data obtained through simulation and the experimental data related to the incorporation experiments where the polydispersion effect was observed. From this study resulted the observation of distinct incorporation velocities for each nucleotide. Several hypothesis were put together to justify why the adenine profile was so different from other nucleotides. The enzyme-substrate complex stability was considered to be one of the hypotheses. This fact can be related to the biologic function of TdT due to the presence of a “lariat-like” loop that is supposed to be related to the correct positioning of the incoming nucleotides.

Index

Figure Index

List of abbreviations

ATP	Adenosine 5'-triphosphate
CTP	Cytidine 5'-triphosphate
dATP	Deoxyadenosine 5'-triphosphate
dCTP	Deoxycytidine 5'-triphosphate
DEACM	[7-(diethylamino)coumarin-4-yl]methyl
DEACM-ATP	P3-[7-(dimethylamino)coumarin-4-yl]methyl adenosine 5'-triphosphate
DEACM-OH	7-diethylamino-4-hydroxymethylcoumarin
dGTP	Deoxyguanosine 5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
dTTP	Thymidine 5'-triphosphate
GTP	Guanosine 5'-triphosphate
HPLC	High-performance liquid chromatography
NMP	Nucleoside monophosphate
NTP	Nucleoside triphosphate
RNA	Ribonucleic acid
TdT	Terminal deoxynucleotidyl transferase
UTP	Uridine 5'-triphosphate

1. Introduction

Nucleic acid synthesis is a vital process to ensure cell viability and transmission of genetic information to future generations. Polymerases play a key role in this synthesis, as they are responsible for numerous biological processes and their forms of regulation.[1] These enzymes may be inserted into two different categories, depending on the substrate: DNA polymerases incorporate deoxyribonucleotides and RNA polymerases incorporate ribonucleotides. DNA polymerases play a vital role in a multitude of processes such as DNA replication, repair and recombination. RNA polymerases, on the other hand, are mainly responsible for the process of transcription, in which a new RNA chain (messenger RNA) is generated from a DNA template. This is the first step for protein synthesis.

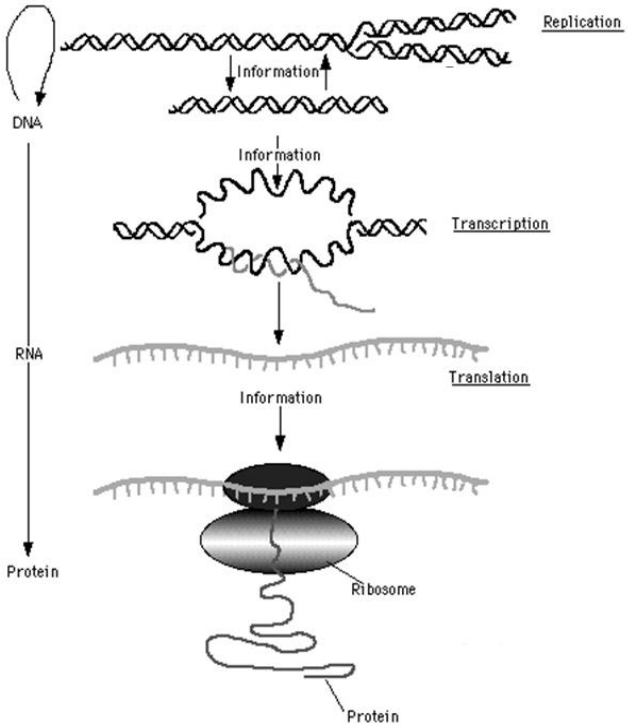

Fig. 1 – Representation of the Central Dogma of Molecular Biology.

In vivo nucleic acid synthesis is a complex and multifunctional process, requiring several key components and strict control mechanisms. However, *in vitro* nucleic acid synthesis is also possible, as one may reproduce optimum conditions for this reaction. Molecular biology techniques such as polymerase chain reaction (PCR),[2] sequencing[3] and *in vitro* transcription[4] are widely used to promote a relatively fast and accurate nucleic acid chain

elongation. Regarding the described techniques, for a successful approach, a DNA or RNA polymerase is always required. This enzyme is often template-dependent as it first recognizes and binds an existing DNA chain. Upon binding, it synthesizes the new chain following the complementary sequence to the previously recognized one through Watson-Crick base pairing. Moreover, the enzyme also demands a short oligonucleotide complementary to the template (primer), to which it binds, and subsequently elongates its 3'-hydroxyl end, using nucleotides as building blocks.[5]

Currently, most oligonucleotides used in molecular biology laboratories are obtained using chemical synthesis protocols.[6] However, the creation of synthetic oligonucleotides *in vitro* could also be possible through the development of an adequate system using enzymatic tools as DNA biological machines. This approach, which can lead to gene synthesis, would be considered a powerful tool for boosting current molecular biology techniques. In order to succeed, a stringent control over the incorporated nucleotides sequence would be necessary. Moreover, the enzyme would have to preferentially possess template-independent activity as to obtain a desired sequence. Combining all these characteristics would result in an adequate system to synthesize gene fragments or even complete gene sequences for downstream biotechnological applications.

Due to its properties, Terminal Deoxynucleotidyl Transferase (TdT) is a unique DNA polymerase which can be a valuable component for the previously described system. This enzyme has the capability to randomly incorporate nucleotides in the 3'-OH end of a nucleic acid chain without the presence of a template strand. Therefore is considered to be a template-independent polymerase.[7] It has also been previously described the unique capacity of TdT to perform, under specific conditions, *de novo* synthesis.[8] In order to efficiently synthesize short oligonucleotide sequences recurring to TdT, a control of nucleotide incorporation in a growing chain is necessary. Therefore, the use of light-induced molecules as nucleotide "cages" has been demonstrated to be an interesting mechanism to achieve such level of control.[9] The general concept is rather simple, in theory, and relies on the idea of using four different derivatives of 4-methylcoumarin to inactivate each nucleotide. Upon a pulse of light of a specific wavelength, the bond

between them would be photocleaved, turning the nucleotide available for incorporation by TdT.

With high concentrations of nucleotides, the polymerization reaction catalyzed by TdT does not present a specific profile of incorporation for each nucleotide, because a distribution of different product sizes has been observed.[10, 11] Therefore it is of vital importance to determine the ideal enzymatic conditions for a controlled oligonucleotide synthesis, i.e. optimum substrate concentration, duration of polymerization, or co-factor selection. Previous DNA polymerases studies have demonstrated that, regarding kinetics, the incorporation of nucleotides, at a single-molecule point of view, follows a stochastic model of probability.[12] This model, which is related to different enzyme processivities, can also be applied to TdT. This assumption leads to the birth of a kinetic computational model that describes the random incorporation of nucleotides by TdT. After proper validation with experimental results, this model can be used to predict the optimum conditions for a specific oligonucleotide chain synthesis.

The main goal of this project is to validate the proposed computational model, based on single molecule kinetics, in order to predict the ideal conditions to an appropriate nucleotide synthesis without the formation of secondary products. Therefore, several reaction conditions need to be tested in order to comprehend how the system works and how can it be changed towards the main purpose of this project. Before attempting this, it is important to analyze some theoretical aspects that fundament this work. Therefore, structural and functional differences between template-dependent and template-independent polymerases will be discussed, as well as their processivities in enzymatic polymerization. Then, light activation of polymerization recurring to 4-methylcoumarin derivatives will be considered in order to ensure an accurate understanding about the system used to achieve control of nucleotide incorporation. Finally, a statistics computational model to predict TdT's kinetic mechanism of nucleotide incorporation on a short oligonucleotide initiator will also be described as well as a brief explanation about single molecule kinetics.

1.1 Nucleic Acid Polymerases

As previously mentioned, DNA and RNA polymerases are fundamental enzymes that are present in all living organisms. These multifunctional enzymes regulate nucleic acid synthesis through dynamic interactions with protein complexes.[1] In fact, evolution itself would not be possible without genome replication and protein synthesis. In order to efficiently elongate a DNA chain, a short oligonucleotide initiator (primer) and the recognition of a template sequence are of vital importance. The primer provides a free 3´-OH end where the incoming nucleotide is incorporated, while the template serves as a base-pairing guideline to a correct polymerization. On the other hand, DNA repair mechanisms, mainly double strand break repair, require the intervention of template-independent polymerases.[5] Several template-independent polymerases have been characterized to date.

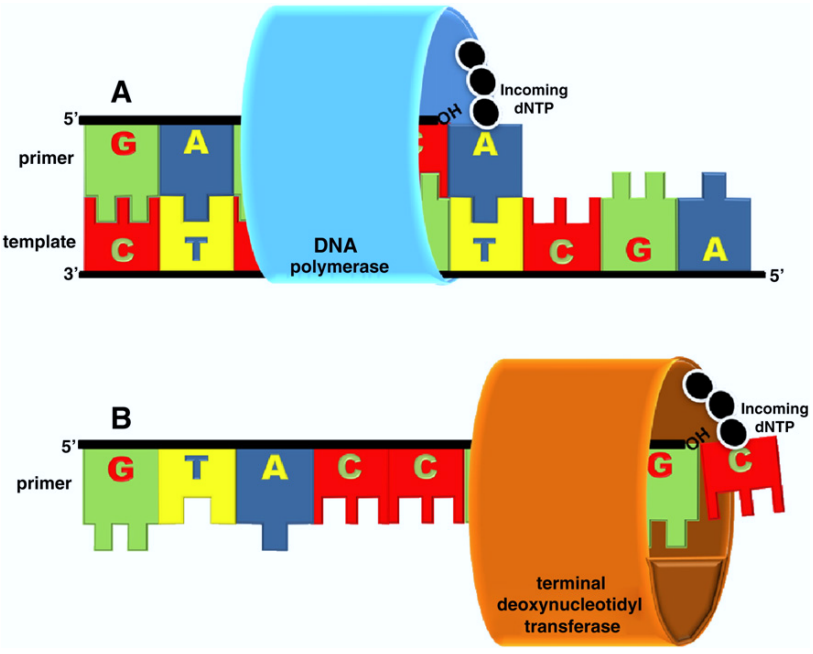

Fig. 2 – Simple model of polymeric strategies for template-dependent (above) and template-independent (below) DNA polymerases. Adapted from Motea *et al*. 2010 [11].

1.1.1 DNA Polymerases classification

An increasing complexity in polymerases' structure and function has been observed from bacteria to superior organisms. Despite this fact, the active site of these enzymes has been highly conserved throughout evolution. Small differences among them made possible the classification of DNA polymerases into six different families (A, B, C, D, Y and X). Each family comprises unique polymerases with specific roles maintaining cellular integrity. However, in this particular project we are interested in template-independent polymerases that belong to Family X of DNA polymerases. Besides TdT, DNA polimerases μ, β and λ are also representative members of this family. Their main roles include repairing mechanisms and bypassing DNA lesions. Here a duplex DNA is always required in order to efficiently assist catalysis.[13]

An important aspect to mention is that TdT is the only template-independent DNA polymerase that catalyzes the incorporation of nucleotides on a single-strand DNA initiator. It is found on primitive lymphocytes and its main role is related to the random addition of nucleotides to a DNA chain during V(D)J recombination. This mechanism of genetic recombination results in an increase of the natural variability of antigen receptors. Besides template-independent activity, TdT has been demonstrated to be extremely inefficient in the presence of double-stranded DNA (dsDNA). This is mainly due to specific structural characteristics namely the presence of a "lariat-like" loop, which behaves like a physical barrier to the accommodation of a template DNA. As such, TdT is a suitable option for extending single-stranded oligonucleotides, but inefficient when in presence of dsDNA. Moreover, the ability to copy a DNA template is completely absent in this enzyme.[13]

1.1.2 Structural characteristics of DNA Polymerases

In order to ensure a correct nucleic acid synthesis, it is important to understand how the enzymatic polymerization is achieved. In this section some structural characteristics that distinguish template-dependent and template-independent polymerases will be discussed. To begin with, all DNA polymerases have similar structures separated into distinct

domains that resemble a right hand's morphology, first proposed for *E. coli* DNA polymerase I, a template-dependent polymerase, in 1993.[14] In the case of template-dependent polymerases, this assembly allows the correct positioning of both primer and template DNA in order to promote the correct nucleotide incorporation. The active catalytic site with important conserved motives, responsible for the phosphoryl transfer reaction, is located in the "palm" domain. The "fingers" domain is responsible for the correct positioning of both the DNA template and the incoming nucleotide at the enzyme's active site. The ability of an enzyme to perform the same reaction on a substrate before the dissociation occurs – processivity, is possible due to the "thumb" domain. This domain binds the duplex DNA when it exits the active site of the enzyme.[11]

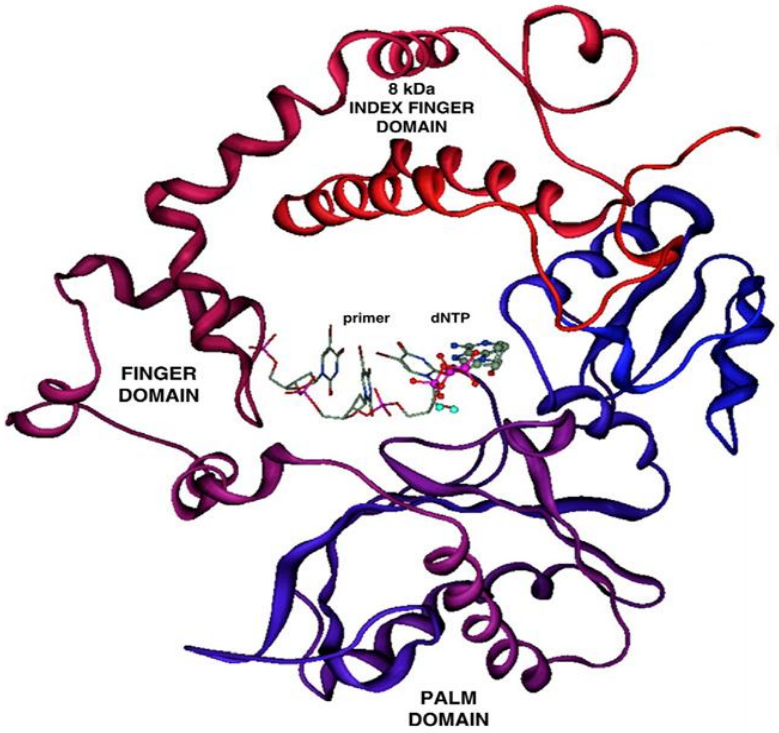

Fig. 3 – Cristal structure of Terminal deoxynucleotidyl Transferase with identification of finger and palm domains. Adapted from Motea *et al*. 2010 [11].

Template-independent DNA polymerases not only share similar morphology features with other DNA polymerases, but also present unique features that define their polymeric activity. TdT is no exception. First, there is a steric hindrance at the catalytic site due to the presence of a unique "lariat-like" loop that avoids interaction with a template DNA.

Moreover, there is an “index finger” domain that acts together with the “thumb” domain to facilitate the diffusion of the incoming nucleotides through the formation of a hole, thus conferring a doughnut-like shape to this enzyme. Direct observation of TdT’s crystal structure indicates that a conformational change might not be necessary in order to assist the polymerization reaction because this enzyme naturally presents a closed like conformation similar to pol β, a family X DNA polymerase with a well-known crystal structure.[15]

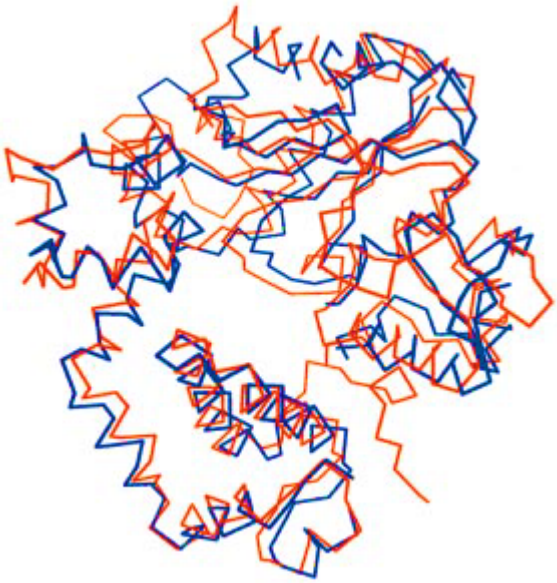

Fig. 4 – Structural overlap of TdT (red) and pol β closed like conformations. Adapted from Delarue *et al.* 2002 [15].

The unique features of TdT presented here made this enzyme an attractive biomolecule to be used in a multitude of applications, ranging from oligonucleotide-end labeling,[16] signaling of double-strand break sites,[11] anti-cancer targeting,[17] to its use as a biochemical tool for downstream approaches.[18]

1.2 Catalytic mechanisms for nucleic acids synthesis

Nucleic acid synthesis, for most DNA polymerases, involves binding of a triphosphate nucleotide to a short pre-existing chain that forms Watson-Crick base pairing with a template sequence. This is possible when polymerases catalyze a nucleophilic attack by the

3'-hydroxyl end of the primer on the dNTP α-phosphate with subsequent release of pyrophosphate. To facilitate the incorporation of an incoming nucleotide, several conformational changes occur within the structure of the polymerase. To begin with, an open conformation of the enzyme is essential to expose the template strand to both the complementary primer and the incoming nucleotide, in order to create an active catalytic site. Upon nucleotide binding, a rotation of finger and thumb domains towards the palm domain takes place, changing the enzyme to a closed conformation. After the incorporation, the growing chain slides unidirectionally across the thumb domain, due to the energy released upon the nucleotide binding, leaving an empty site behind. Following a new conformational change and the release of pyrophosphate (PPi), the enzyme's active site is available for the incorporation of the subsequent nucleotide. This is a cyclic process which allows the incorporation of a single nucleotide at a time.[19] So far only a general model has been described. The scheme presented in figure 5, illustrates this model for DNA template-dependent polymerase kinetics. Template-independent polymerases are thought to follow a similar mechanism, although more work is necessary to confirm this assumption and to determine the specific polymerization rate constants.

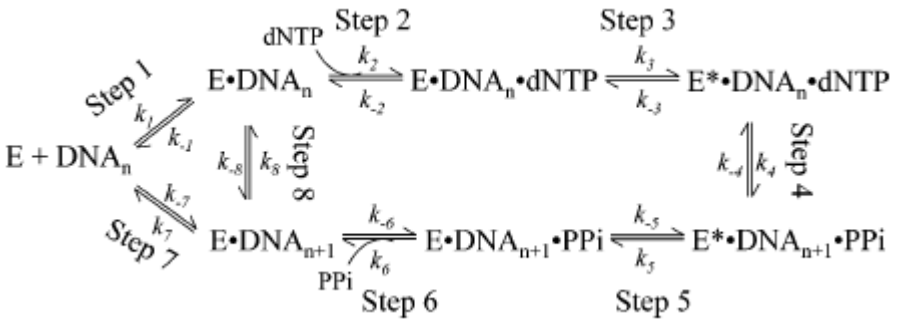

Fig. 5 – Schematic representation of the general model described for DNA template-dependent polymerase kinetics. A strictly ordered polymerase recognition of template DNA is followed by binding of the incoming nucleotide. Then, a conformational alteration takes place and the phosphoryl transfer reaction is achieved. PPi is released as the enzyme acquires the original conformation leading to the reaction's progression. In Fowler *et al*. 2006 [13].

If the polymerase is said to be processive, when a new cycle of incorporation begins, the translocation of the enzyme through the template chain without dissociation becomes a possibility. However, information derived from TdT's crystal structure led to the belief that

template-independent polymerases act in a strictly distributive manner upon primer binding. This means that it probably dissociates from DNA after each incorporation cycle. [13] Therefore it appears the reaction is slower than expected due to cycles of binding and releasing of extended nucleic acid chains.

On template-dependent polymerases, the polymerization reaction requires the presence of two divalent ions, usually Mg^{2+}, on the enzyme's active site. These metal ions are coordinated by three carboxylate residues (aspartate or glutamate), highly conserved among all DNA polymerases families. The deprotonation (lower pKa) of the primer's 3'-OH end, a crucial step for the reaction to occur, is assisted by one Mg^{2+}, allowing the nucleophilic attack of this group to the α-phosphate of the free nucleotide. The pentacovalent transition state of the α-phosphate is stabilized by the other Mg^{2+} allowing the release of pyrophosphate. Template-independent polymerases also require divalent metal ions for DNA catalysis, but in contrast to template-dependent ones, a variety of cations such as Mg^{2+}, Co^{2+}, Mn^{2+} and Zn^{2+} can be used. TdT presents itself as a good example, as its catalytic efficiency changes according to which divalent cation is present. For example, small amounts of Zn^{2+} combined with Mg^{2+} have been shown to increase the polymerization rate of all nucleotides, while Mn^{2+} alone is less efficient. It has been proposed that each metal ion influences the enzyme's nucleotide incorporation kinetics. According to Chang and co-workers, the preferential incorporation of dGTP and dATP is achieved when Mg^{2+} is used. On the other hand, Co^{2+} increases the incorporation of dCTP and dTTP.[11, 13, 19]

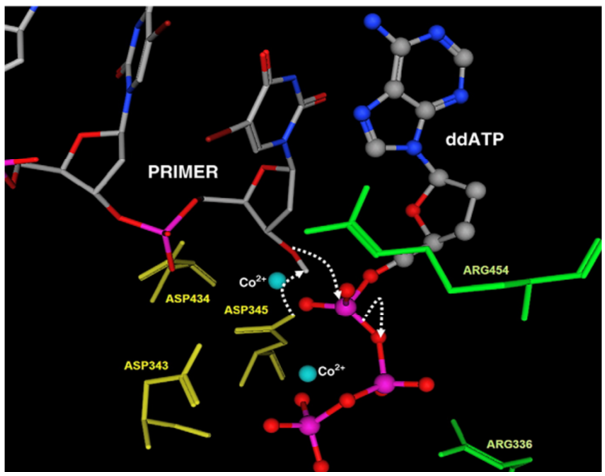

Fig. 6 – A structural overview on the active site of TdT. Highlights on aspartate residues coordinating cobalt divalent ions. In Motea *et al*. 2010 [11].

1.2.1 Kinetic mechanisms

In the same way as other DNA polymerases, TdT's polymerization reaction, with a primer ((pdN)n) and dNTP as substrates, may be described as:

$$(pdN)_{n\geq 3} + dNTP \leftrightarrow (pdN)_{n\geq 3}pdN + PPi$$

As we can observe, a trinucleotide primer with a minimum length of three nucleotides, and bearing a free 3'-OH and 5'-phosphate ends, is frequently required.

In template-dependent polymerases a strict order related to the recognition and binding of dNTP and DNA substrates to the enzyme's catalytic core is essential. Because of a sequence specific dependence and the need for correct nucleotide incorporation, the binding of the incoming dNTP takes place after the recognition of the template DNA that is being copied. Template-independent polymerases such as TdT, on the other hand, follow a random mechanism for polymerase recognition of its substrates. Such observations were possible after kinetic studies with some inhibitor molecules.[11] Moreover, TdT shows different affinity for each nucleotide. It seems that dGTP and dTTP have a higher incorporation rate, when compared with dATP, probably due to the stronger hydrogen-bonding that is formed during polymerization.[11] Also remarkable is the ability of TdT to incorporate nucleotide analogues like 5-substituted indolyl deoxynucleotides, 2',3'-dideoxynucleotides (ddNTPs) or α-D-dNTPs. Because of these findings, it was proposed that the nucleotide binding would not depend entirely on the interactions between the positively-charged amino acids on the catalytic center of the enzyme and the negatively-charged triphosphate nucleotides. Therefore, TdT is the ideal candidate for this project. Besides nucleotide recognition and incorporation, the new chain elongation is also related to the nucleotide present in the 3'-OH end of the growing chain. Some analogues do not allow efficient elongation rates when compared with natural nucleotides.[11]

1.3 Nucleotide selectivity by DNA Polymerases

Interestingly, i*n vitro* published experiments also demonstrated the ability of TdT to incorporate ribonucleotides (NTPs) into a pre-existing chain. However, TdT's preference is approximately 8-fold superior in the case of deoxynucleotides (dNTPs). In order to understand the mechanism beyond the selectivity towards dNTPs we need to look how polymerases evolution contributed to this achievement. Such level of selectivity has been developed by DNA polymerases to restrict the misincorporation of NTP substrates in DNA synthesis when the ratio NTP/dNTP is relatively high. Several mechanisms to discriminate between dNTPs and NTPs have been described.[20] Some of those mechanisms will be presented next.

To begin with, most DNA polymerases developed a simple steric exclusion mechanism in which a residue present on the enzyme's active site recognizes the 2'-OH group on the ribose ring of an incoming NTP. This recognition is based mainly in size and stereochemistry of the 2'-position group. As a consequence, there is a slower rate of incorporation. Polymerases with proofreading activity, the capacity to correct some mistakes during the polymerization reaction, take the advantage of slower activity on the catalytic active site. This allows the activation of the exonuclease active site and the removal of the incorporated NTP. Besides the slower rate of NTP incorporation by common DNA polymerases, extensions of primers containing NMPs often leads to premature termination of the primer elongation to approximately four incorporations. Again is proposed as a consequence of an inefficient alignment of the 3'-OH end, the presence of the 2′-OH on the ribose ring which also prevents the stabilization of the template-primer hybrid.[20]

1.4 Stochastic models to predict TdT's enzymatic progression

One of the main goals of this project is to derive an expression for the extension rate of DNA polymerases, in particular TdT. An outcome of the proposed model is that it could predict extension rate, fidelity, processivity and product yield in *in vitro* nucleic acids

synthesis. In order to understand the mechanism by which enzymes, in particular DNA polymerases catalyze a specific reaction, an extensive study on the enzymes kinetic mechanism is essential.

The classic method for characterization of enzyme activities is based on the work developed by Michaelis and Menten. It describes, through chemical equations, the reversible binding of a substrate (*S*) to an enzyme's (*E*) catalytic center, to form an enzyme-substrate complex (*ES*). The reaction product (*P*) is formed and the original enzyme conformation (*E*) is generated from E^0.[21, 22]

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E^0 + P, \qquad E^0 \xrightarrow{k_3} E \tag{1}$$

There is a hyperbolic dependence of the product formation rate (v) on the substrate concentration denoted by the classic Michaelis-Menten equation, where v_{max} is the maximum enzyme velocity, or, in other words, the reaction velocity at saturating substrate concentration [*S*]. The Michaelis-Menten constant, substrate concentration at which the enzymatic velocity is half of v_{max}, is defined as $K_M = (k_{-1} + k_2)/k_1$.[21, 22]

$$v = \frac{v_{max}[S]}{[S]+K_M} \tag{2}$$

This method accurately describes the enzymatic kinetics for large ensemble of enzyme molecules. However, previous studies also reported the use of Michaelis-Menten equations for single-molecule approaches. Enzymatic studies that rely on single-molecule kinetics take into account the contribution of enzyme´s intermediates and dynamic disorders that are not reported by classic equations. These disorders are direct consequences of the fluctuations in the enzymatic rates associated with conformational variations.[21, 22]

1.4.1 Single-molecule kinetics

In single-molecule experiments the progression of enzymatic reactions can be considered a stochastic event, or that is directly dependent on random events. In other words, it focuses on the measurement of the probability density of stochastic waiting times for each enzymatic turnover (kinetic cycle). To record these turnovers, each enzyme molecule is monitored constantly over time as it experiences all different conformations from *E* to *ES* and to E^0. The probability density function is therefore obtained by collecting all the information available about the previously recorded enzyme turnovers.[23]

In classic kinetics, mainly deterministic, the enzymatic rate constants are obtained recurring to molecular concentration variations. This approach represents the time evolution as a continuous process directed by coupled first-order differential equations. On the other hand, single-molecule kinetics is formulated in terms of the probabilities about where a single enzyme is on a particular reaction state at a specific time. In this case, the time evolution is considered a random process ruled by one differential equation, the Chemical Master Equation (CME). This equation is simply a mathematical form of describing the systems evolution considering its initial conditions. With this in mind, we might consider that the rate of product formation is no longer characterized by one rate constant related to the enzyme, but instead by more than one. Because of that, a deterministic approach of the system´s evolution is not enough to forecast its future. Especially in biological systems, where some molecular populations are small, a stochastic approach can be very efficient.[21] In fact, the time evolution of chemical systems cannot be truly described by continuous nor deterministic methodologies. First, the molecular population changes are discontinuous throughout the entire reaction time, as they change only by integer amounts. Then, regarding the prediction of the evolution of molecular population values is fundamental to consider the exact positions and velocities of all the system´s components, so it is not deterministic either. In order to apply this methodology to DNA polymerases systems a simple explanation of stochastic chemical kinetics will be presented here.[21]

An example of TdT's reaction system will be discussed. It is important to consider a homogenous system in a constant volume *V,* and to admit that is in thermal (not necessarily chemical) equilibrium at a specific temperature *T*. The main question here is: considering each molecule species and their quantities present in the system in the beginning of the reaction, what will be the system's distribution in terms of molecule populations in a later time? In other words, the main goal is to estimate the state vector $X(t) = (X_1(t),\ldots,X_N(t))$, given that the system was in state $X(t_0) = x_0$ in the initial time t_0. A chemical reaction, in order to occur, requires a collision between two or more species. These collisions can be described to occur in a random manner because the system is in thermal equilibrium. As an example, considering two molecules S_1 and S_2 with radius r_1 and r_2 they will collide only when the $r_{12} = r_1 + r_2$ condition is verified.[24] To understand the evolution of a reaction channel is important to consider a state-change vector $v_j = (v_{1j},\ldots,v_{Nj})$, as a variation in the molecular populations caused by a reaction event. Moreover, there is a need to define a propensity function *aj* as

$$aj(x)dt \triangleq \text{the probability, given } X(t) = x, \text{ that one } R_j \text{ reaction will occur somewhere inside } V \text{ in the next infinitesimal time interval } [t, t + dt]. \quad (3)$$

[24]

Each propensity value $a(x)$ is related to the reaction velocity in a way that is characterized by the enzyme rate constant and the initial conditions of the molecule populations at a particular time. The evolution of these values represents an accurate description of the reaction state changes. Here is considered the random addition of nucleotides into an oligonucleotide chain by TdT described in a simple way:

$$E + I + N \xrightarrow{k} E + IN. \quad (4)$$

The reaction proceeds until the established end-time is achieved or until the nucleotide concentration reaches a critical value.

In the previous examples the existence of an enzymatic constant *k* that accounts for TdT's activity is conceptual. This enzymatic rate constant is considered to remain unaltered throughout the reaction.[24] With this in mind, is possible to calculate the reaction velocity *v*,

according to the propensity function, considering k and the initial concentrations of initiator I and nucleotide N molecules as follows

$$v = k[I][N]. \tag{5}$$

Then, following a deterministic approach accurately applied to a system with a reasonable number of molecules, after an infinitesimal time evolution dt, is possible to determine the next reaction velocity considering the molecular changes in the system. So, with

$$E + IN + N \xrightarrow{k} E + IN_2, \tag{6}$$

we can calculate the enzymatic reaction velocity as

$$v = k[IN][N]. \tag{7}$$

The system's evolution is monitored following the same patterns leading to the calculation of a reaction velocity for each nucleotide incorporation cycle.

The stochastic approach is generally used in biochemical applications that involve the study of enzymology.[23] Its fundamental equation, CME, is derived from the "probability density function" $P(x, t \mid x_0, t_0)$ that evolves through a "jump Markov process" to $P(x, t + dt \mid x_0, t_0)$. This means that the system's evolution depends on the initial molecular values and changes discontinuously with time. According to the Tau Leap algorithm, each time jump evolves depending on the Poisson probability distribution. Due to an increased computational complexity related to solving the CME, the common adopted solution is to use Monte Carlo algorithms to track the time evolution of the system. An additional feature related to the computational analysis of enzyme kinetics is the definition of a restriction condition that leads to the end of the CME. Often, each tau leap "jumping time" is limited to a pre-defined value

$$P(x, t|x_0, t_0) \triangleq Prob\,\{X(t) = x, given\ X(t_0) = x_0\}.^{24} \tag{8}$$

The evolution of the probability function includes the probability that the system is unaltered in (t + dt), because it is already on that state, and the systems evolution due to one reaction in (t + dt) as follows.

$$P(x, t + dt|x0, t0) =$$

$$= P(x, t|x0, t0) \times [1 - \sum_{j=1}^{M} (aj(x)dt)] + \sum_{j=1}^{M} P(x - vj, t|x0, t0) \times (aj(x - vj)dt) \quad (9)$$

A deterministic reaction rate naturally expresses the average number of reactions taking place in the biological system per unit time. For stochastic enzyme kinetics, when applied to a large number of molecules, the propensity function $aj(x)$dt can be found in the mean value of the *poisson* probability distribution at [t, t + dt]. Therefore, with a large number of molecules, both deterministic and stochastic approaches present the same prediction of the system's evolution.[21]

Now is time to consider the simulation algorithm that allows the prediction of the system's evolution using the Chemical Master Equation. Essentially, this is a simulation of the time evolution of the reaction system. The main goal here is to predict when the next reaction will take place and what kind of reaction will it be. A sequence of simple computational algorithm steps used will be presented next.

Step 1: First, there is a need to define all the desirable initial variables, in order initialize the reaction. Is defined the duration of the reaction by t_0 and t_{END}. Moreover, there is a need to set the initial concentrations of the population molecules (substrates), a short oligonucleotide initiator and nucleotides.

Step 2: The propensity function $aj(x)$ is calculated for a particular time interval considering a specific rate constant k related, as described above, to the enzymes activity as follows:

$$aj(x) = k \times x(i) \times x(i + 1). \quad (10)$$

Step 3: The next step is to generate random numbers from the distributions with probability density/mass functions $P(x, t)$ and to generate random τ time jumps.

Step 4: Finally the system is updated to a new state where $t = t + dt$ and $x = x + dx$. The incorporation cycle is then finished. A repetition from step 2 is performed until the end of the reaction (t_{END}) is reached.

How the system evolves through the Tau-leaping method is an interesting way for accounting many reactions with only one time jump. Its main purpose is to avoid situations where the time evolution $[t; t + dt]$ of the system is not enough to cause an expected change in the molecular populations. As a consequence there is no revealing alteration on the propensity function values and the system is said not to evolve. Therefore, there is a need to define a Poisson random variable that represents the reaction mixture time evolution and satisfy the leap condition here described. Within each polymerization cycle the time interval is randomly generated and the propensity function is updated according to the observed molecular changes.[24]

1.5 Light activation of nucleic acid synthesis

The discovery of how enzymes, particularly DNA polymerases, work and how they are regulated is crucial to ensure an efficient manipulation of *in vitro* nucleic acid synthesis. In order to regulate a polymerization reaction, a strict control of one of the main components of this synthesis, namely the polymerase, the initiator oligonucleotide or the nucleotides, is required. Our group has described some work focused on control of the nucleotide availability during enzymatic reactions. Due to previously obtained results in this area, we targeted these molecules, and opted to use caging compounds that are activated through light in order to control externally the course of the reaction. Light has been used in fast activation or deactivation of biological functions such as gene expression and consequently protein synthesis.[25] A major advantage of this approach is related to the fact that light represents a distinctive non-reactive tool that can be controlled in amplitude, pulse timing and incident area.[26]

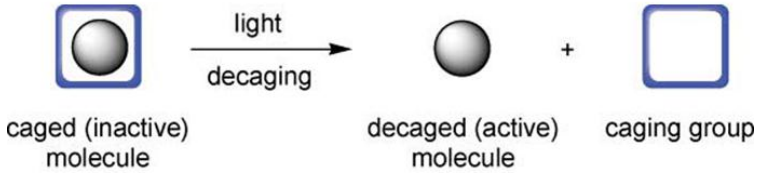

Fig. 7 – Schematic representation of the process of light-induced release of a caged substrate. Adapted from Young *et al*. 2007 [26].

Nucleotide incorporation studies depend not only on the enzymatic activity but also on the enzyme's ability of discrimination between each one of the four nucleotides. Therefore, if such control of incorporation is absent, a normal distribution of nucleotides in a growing chain would be observed. Moreover, even with a small control of the elongation enzymatic step, the same would probably happen. In enzymatic primer synthesis this scenario would be unwanted, especially with the error propagation derived from nucleotide incorporation. In fact, it is of vital importance to ensure a specific oligonucleotide sequence derived from ordered nucleotide incorporation.

The control of enzymatic polymerization became possible after Kaplan and coworkers first introduced the term "caging" and applied it to biological applications.[27] They used a nitrobenzyl protecting group to cage an ATP molecule. Such protection effect could be removed upon irradiation with UV light.[27] This level of control is achieved through light when derivatives of 4-methylcoumarin are used as protection agents for nucleotides. These photosensitive molecules are excellent candidates to be used in biological molecules synthesis, due to their low reactivity. Upon light stimulus of a determined wavelength, the bond between the caging molecule and the nucleotide is destroyed and the nucleotide becomes available for incorporation into the growing chain, by the polymerase.[28]

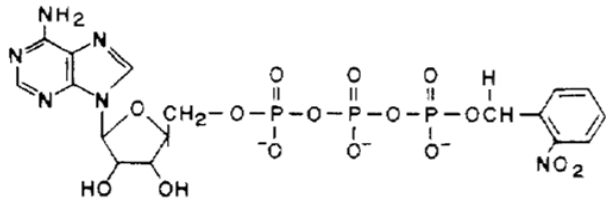

Fig. 8 – First nucleotide ever caged. Here is presented the binding of 2-nitrobenzyl to ATP

A multitude of protecting agents has been developed in order to inactivate several compounds. Divalent ions, e.g. calcium, small proteins, neurotransmitters and nucleotides are examples of such molecules.[28] Previous attempts to control nucleic acid synthesis, whether in transcription with RNA polymerase or DNA synthesis with Klenow polymerase have produced valuable results. However, such strategies led to the premature termination of template elongation steps, resulting in a permanent end of the nucleic acid synthesis.[29,30] Therefore, nucleotide caging would constitute a more adequate alternative when attempting to control nucleic acid chain elongation. Some nucleotide caging photoprotective groups, due to their relevance to the presented work, will be described next.

In order to achieve such level of control, coumarin derivatives were used in preliminary studies to functionalize cAMP and cGMP through direct interactions with the phosphate groups.[31] This strategy is believed to cause polymerization rate alterations. To achieve an efficient control of the incorporated nucleotide, four photoprotective groups with different absorption wavelengths are needed. Although other molecules have also been used as nucleotide cages, main attention falls on coumarin derivatives, due to its interesting characteristics. A noteworthy, yet important feature of coumarines is the ability to absorb light near the visible region of the absorption spectrum (<420 nm).[32] Another coumarin derivative molecule that is important to highlight is [7-(diethylamino)coumarin-4-yl]methyl (DEACM) and 4-(hydroxymethyl)-7-methoxycoumarin (MCM). Whilst DEACM presents high extinction coefficient and photochemical quantum yield, MCM exhibits lower extinction coefficient and photochemical quantum yield, but it is very stable upon hydrolysis. It is extremely important that these two molecules, present a considerable absorption near the visible region in order to avoid potential damaging on DNA from UV radiation when light is used as an external release stimulus for the cage molecule, towards application to biological systems.[25,32]

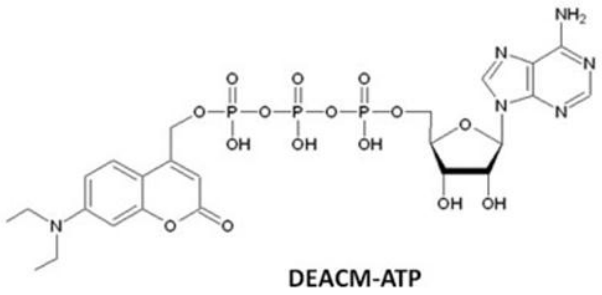

Fig. 9 – ATP molecule caged with DEACM. Adapted from Pinheiro *et al*. 2008 [32].

The use of coumarin-caged nucleotides as building blocks for controlled *in vitro* transcription was reported by Vidal Pinheiro and co-workers.[32] This elegant approach relies on the use of ATP molecules caged with DEACM (DEACM-ATP) for light-controlled activation of *in vitro* transcription reaction by the RNA polymerase from phage T7. With caged ATP there is no RNA synthesis since T7 RNA polymerase is unable to perform chain elongation, and consequently transcription. Upon DEACM-ATP irradiation with 390 nm light, the ester bond between the ATP molecule and the caging molecule is cleaved. After photocleavage, the ATP becomes available for incorporation and the *in vitro* transcription reaction proceeds, with the formation of 7-diethylamino-4-hydroxymethylcoumarin (DEACM-OH) as a secondary product of the photocleavage.[32] The same mechanism described above will now be used for light-controlled activation of nucleic acid polymerization reaction by TdT DNA polymerase.

2. Materials and Methods

All chemicals were purchased from Sigma Aldrich of the highest purity available. The 20-mer oligonucleotide Abl_Fw (5'- cgg tgg ccg acg ggc tca tc - 3') used as initiator of the polymeric reaction was purchased from STAB Vida, Portugal. Terminal deoxynucleotidyl Transferase, reaction buffer Y-Tango and free Nucleotide Triphosphates (dNTPs and NTPs) were purchased from Fermentas, USA.

2.1 Nucleic acids synthesis with Terminal deoxynucleotidyl Transferase

All enzymatic reactions (final volume of 20 µL) were carried out in triplicate using 1x Y-Tango buffer (33 mM Tris-acetate, 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/mL BSA; pH 7.9), 1 mM cobalt chloride, 25 pmol of the 20-mer initiator (Abl_Fw), 20 U of Terminal deoxynucleotidyl Transferase and free nucleotides. Different ratios of Initiator:Nucleotide (I:N; 1:1; 1:2; 1:3) were used with both dNTPs and NTPs to evaluate the progression of the reaction.

The reaction was carried out on a MyCycler™ Thermal Cycler (Bio-Rad, USA) at 37ºC for 60 minutes, followed by the 1:1 (equimolar) addition of formamide solution (1 mL of formamide, 20 uL of EDTA 0.1M, pH 8) and 15 minutes at 75ºC to inactivate the reaction. The obtained products were separated in denaturing 12% polyacrylamide gel electrophoresis (50 mL volume) at constant temperature (20º C) with 8 M urea at 4 W for 195 minutes in 1x TBE. The results were visualized in a UV transilluminator after post-staining with ethidium bromide (0.5 µg/mL; 20 minutes) followed by 10 minutes washing with distilled H_2O.

The quantification of the product bands was calculated from plots of pixel intensity versus band mobility (rf) using the Quantity One software (Bio-Rad, USA). To calibrate each gel measurement, GeneRuler™ Ultra Low Range DNA Ladder and the initiator itself (Abl_Fw) were used as standards. A calibration curve was then created to correct any differences between subsequent experiences.

2.2 Monitorization of Light-induced decaging of DEACM-ATP

The DEACM-ATP solution used in this project was synthesized according to a previously published method.[32] All decaging experiments were carried out on a SPEX Fluorolog model spectrometer (Jobin Yvon Horiba, Edison, NJ) with 1681 Spex and 0.22 m excitation module (18 nm slit bandwidth). DEACM-ATP (100 µM) irradiation was performed in 60 µL fluorescence quartz cells with increasing time (10, 20, 30, 40, 50, 60 minutes) at 390 nm with an I_0 of 4.97 x 10^{-9} Einstein/cm^3/s.

The irradiation profiles were determined by HPLC through monitorization of the depletion of DEACM-ATP and the formation on DEACM-OH. The DEACM-OH solution (300 µM) was prepared by dissolving 1.48 mg of DEACM-OH in 10 mL of acetone. This solution was then evaporated under vacuum until the formation of a thin film, followed by the addition of 20 mL of hot distilled water and ultra-sound bath until a clear solution was obtained. The HPLC analysis was performed in a Hitachi-Merck HPLC L6200A Pump with an L-4500 Diode Array Detector using a Polystirene-Divinylbenzene (PLRP-S) semi-preparative column from Polymer Labs (4.6 mm x 150 mm, 8 µm, 300 Å). Triethylammonium acetate buffer (5 mM, pH 6.9) was used as eluent A; Methanol was used as eluent B. The gradient started with 70% of A and 30% of B, after 3 minutes there was an increase of B to 90% until 5 minutes and the gradient finished at 10% of A and 90% of B at 12.5 minutes. These analyses were performed at a column temperature of 35º C and at a 2 ml/minute flow rate. The monitorization of the separation and determination of the irradiation profile was performed by UV spectroscopy at 390 nm and peak area quantification.

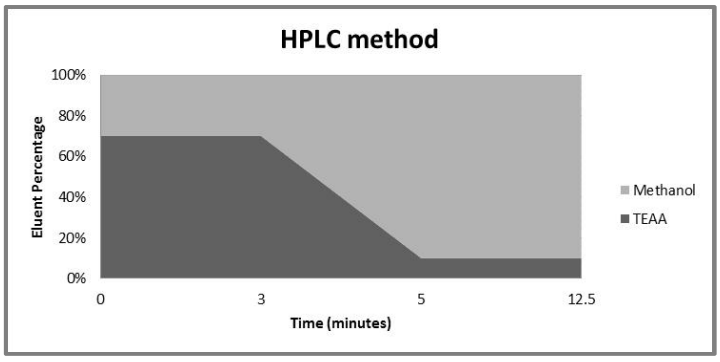

Fig. 10 – Area chart illustration of the HPLC analytic method used to follow the decaging of ATP.

2.3 Terminal deoxynucleotidyl Transferase light activated polymerization

Light activated polymerization was performed as described in section 2.1, using an equivalent amount of DEACM-ATP before and after irradiation at 390 nm (I:N; 1:1, 1:10), as a substitute for free nucleotides. Only after irradiation the DEACM-ATP solution was added to the polymerization reaction to evaluate its effect. An inhibition assay with increasing concentrations of DEACM-OH was also performed. To each sample, where ATP was the free nucleotide source, 10, 25, 50, 75, 100 and 150 µM of DEACM-OH was added and the results were obtained as described above in section 2.1.

2.4 Modulation of Terminal deoxynucleotidyl Transferase polymerization

The percentage of incorporated nucleotides was determined according to data obtained from peak area quantification, for all samples. Afterwards, percentage of incorporation was plotted versus the number of incorporated nucleotides for each single experiment. The obtained data was later fitted into an appropriate distribution function. The probability value that fitted properly with the obtained results was then compared with the data obtained from the mathematic model.

MATLAB (MathWorks) is the software for numerical analysis that allowed the creation of the simulation model, based on DNA polymerases' catalytic function. It begins with some standard inputs: reaction time, initial concentration of nucleotide, initial concentration of initiator, maximum integration interval, and extension of the tau leap jump. Then, the algorithm progresses according to the definition of specific parameters such as E (related to the enzymatic concentration) and k (related to the enzymes catalytic velocity). The algorithm provided by Prof. Rui Oliveira (REQUIMTE/CQFB, FCT/UNL), calculates the probability of an incorporation to occur based on a statistical distribution of probability - Binomial and Poisson, and the output is presented as a bidimensional plot, with the distribution of formed products versus their relative proportions. These results were then adjusted into probabilities and compared with the biological results. After the fitting is performed, a proper value of initial velocity of incorporation (k) can be determined for each nucleotide.

3. Results and Discussion

3.1 *In vitro* polymerization with Terminal deoxynucleotidyl Transferase

Terminal deoxynucleotidyl Transferase was the selected enzyme to perform this study as it presents the ability to synthesize single strand DNA chains without a template sequence, requiring only a short oligonucleotide initiator (primer).[33] It is known, according to previously published work, that TdT incorporates selectively each nucleotide.[11] Regarding the main objectives of this thesis, it was imperative to begin with an experimental design to attain similar results, using previously reported work conditions. In vitro DNA synthesis roughly requires a DNA polymerase, a primer sequence to act as a polymeric initiator and free nucleotides as building blocks. Moreover, it is important that all *in vitro* reactions show reliability throughout time, so in the beginning, every experiment was performed in triplicate with equimolar concentrations of initiator and nucleotide.

Gel electrophoresis is an easy and widely used analytical method in molecular biology labs to visualize and separate short DNA molecules, according to their molecular weight, conformation or net charge. In this project, denaturing 12% polyacrylamide gels were used to guarantee similar conformation and net charge properties amongst all samples while achieving a maximum resolution of the observed bands for downstream applications related to the quantification of the reaction products. After obtaining the desired single-base extension result, an approach based in gel band densitometry was adopted in order to quantify and characterize the enzymatic incorporation of nucleotides. Pixel density is directly proportional to the intensity of fluorescence and may, therefore serve as an indicator to estimate the nucleic acid quantity in each band.[34] Such approach has been used previously to evaluate DNA polimerase's insertion fidelity.[35] The obtained results are highlighted in figure 11, as a plot of relative intensity values *versus* the number of incorporated nucleotides in each case. Despite being simple, this method cannot provide the exact quantification of nucleic acids present in each gel band. Moreover, differences between samples cannot be minimized unless a comparison between each sample and a reference is performed. The reaction initiator and GeneRuler™ Ultra Low Range DNA Ladder were used to standardize the samples obtained from different gels and allow their

comparison. This approach accounts for a relative quantification of the nucleotides incorporated during polymerization reaction, thus giving an insight on its efficiency.

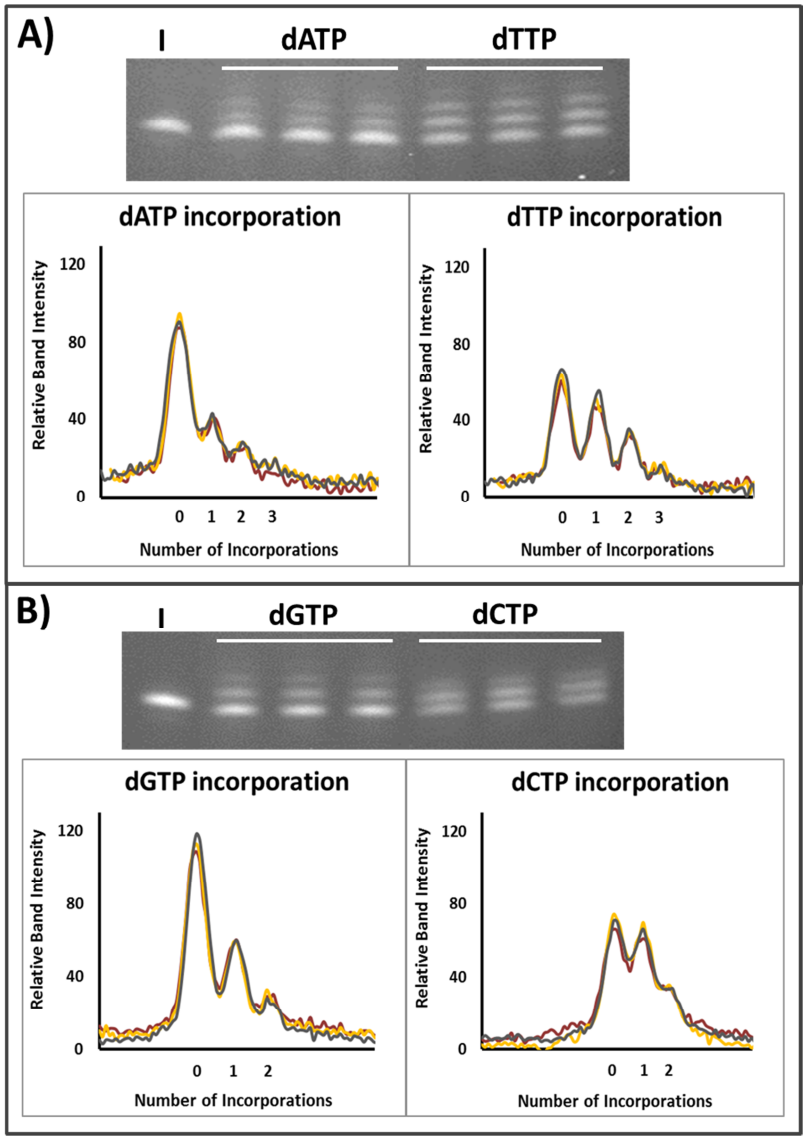

Fig. 11 – Terminal deoxynucleotidyl Transferase (TdT) polymerization reaction where deoxyribonucleotides (dNTPs) are incorporated into a 20-mer oligonucleotide initiator. 12% denaturing polyacrylamide gel image of the incorporation of: dATP and dTTP (**A**), dGTP and dCTP (**B**); gel band density profile for each nucleotide is presented, the left peak represents the primer initiator with 20 nucleotides of length and the subsequent ones are related to sequential additions of one nucleotide at a time to the initiator (21-23 nucleotides).

Figure 11 demonstrates the insertion profile for every deoxynucleotide resultant from a polymerization reaction with TdT and equimolar concentrations of initiator and dNTPs. Each band represents a reaction product with a specific size (20 to 23 nucleotides). From left to right we can observe the initiator molecule (20 nucleotides long) where no nucleotide was inserted by TdT, followed by the reaction products, where addition of 1, 2 and 3 nucleotides respectively, was performed by this enzyme. The intensity of each band indicates the relative amount of each sample present in the mixture. In all experiments we can observe significant amounts of initiator molecule where no incorporation took place, confirming that longer oligonucleotide initiator chains are preferred as targets for future incorporation reactions by TdT, as previously demonstrated.[33] This clearly reveals a difference on the capability demonstrated by TdT for polymerizing each deoxynucleotide into a pre-existing nucleic acid chain. It appears that the incorporation of dTTP, dCTP and dGTP are much more efficient then incorporation of dATP.[11] Such differences might be related to the biological role of TdT. However, this must be considered when adapting the kinetic model to the enzymatic reaction.

3.1.1 Differential incorporation of nucleotides by TdT

Theoretically, if the same number of initiator sequences and the same number of nucleotides were put together in a reaction system, the DNA polymerases present in that system would incorporate only one nucleotide per initiator sequence. In practice however, this does not occur. Instead, a statistical distribution of products resulting from the described system is observed (Fig. 11). Such effect can be explained if we think of each enzyme as an individual organization with different kinetic rates that result in distinct enzymatic processivities. To evaluate this hypothesis, it is important to study the system in its natural form and observe its evolution within time, with increasing concentration of nucleotides.

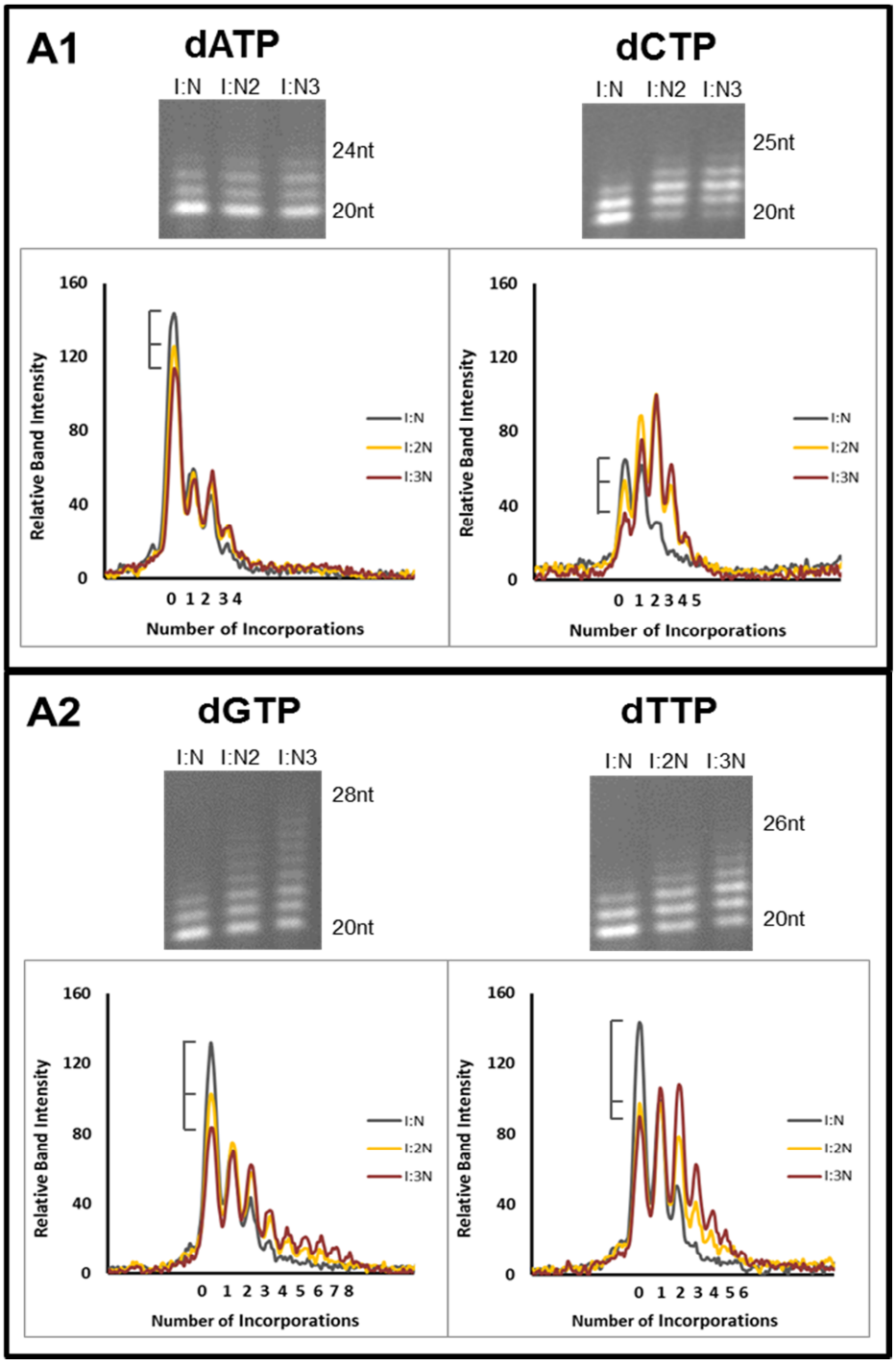

Fig. 12.1 – TdT polymerization reaction with increasing concentration of dNTPs. 12% denaturing polyacrylamide gel image of 1:1, 1:2 and 1:3 incorporation of: dATP and dCTP (A1), dGTP and dTTP (A2); gel band density profile for each nucleotide, where individual peaks from left to right represent different product sizes (20-28 nucleotides); the marker near the 20 nucleotide peak illustrates the consumption of initiator molecules with increasing concentration of dNTPs.

Figure 12.1 shows the electrophoretic patterns and relative band intensities of polymerization reactions using deoxynucleotides (dNTPs) as a substrate. The principle applied to this figure is the same as in figure 11. The plot represents, from left to right, the initiator molecules that did not participate in the reaction, followed by the reaction products resulting from dNTP addition to each incorporation site of the initiator. Increasing the concentration of polymeric building blocks resulted in an impressively different profile for each dNTP. Each distinct profile results from different incorporation velocities. For example, the incorporation profile for dATP shows that few initiator molecules participated in the reaction and no more than 3 nucleotides were incorporated by TdT in significant quantities. This result is an indicator of a slow incorporation reaction when dATP is the only nucleotide source present. On the other hand, there is the incorporation profile of dGTP with the longest products formed, 8 incorporations in total, for a I:N ratio of 1:3. In this case, a high rate of incorporation is probably the cause as TdT seems to prefer, in each cycle of incorporation, initiator molecules that already presented a dGMP incorporation in their 3'-OH end. Regarding the number of incorporations, dCTP and dTTP present intermediate results where the mean value of nucleotide incorporation is 2 in both cases. This result is closer to what was expected with a 1:3 I:N ratio and is also consistent with an intermediate nucleotide incorporation rate. In general, a rather wide size distribution of reaction products is observed, regardless of the incorporated nucleotide.

In order to restrict the obtained polydispersion, a more stringent control over the polymerization reaction has to be achieved. Therefore, there is a need to limit the available substrates to the enzyme active site according to their incorporation velocities. In this case, the use of caged nucleotides seems to be a valid alternative to achieve such level of control. During the PhD project of Vidal Pinheiro, several attempts were made in order to conjugate deoxynucleotides with caging molecules. However, this strategy was only effective when coumarin derivatives were used to cage ribonucleotides (NTPs), particularly ATP and GTP. The system was first introduced in T7 RNA polymerase *in vitro* transcription reaction[32] however, due to the lack of selectivity presented by TdT towards its substrates, this strategy is also feasible to be applied in this project.

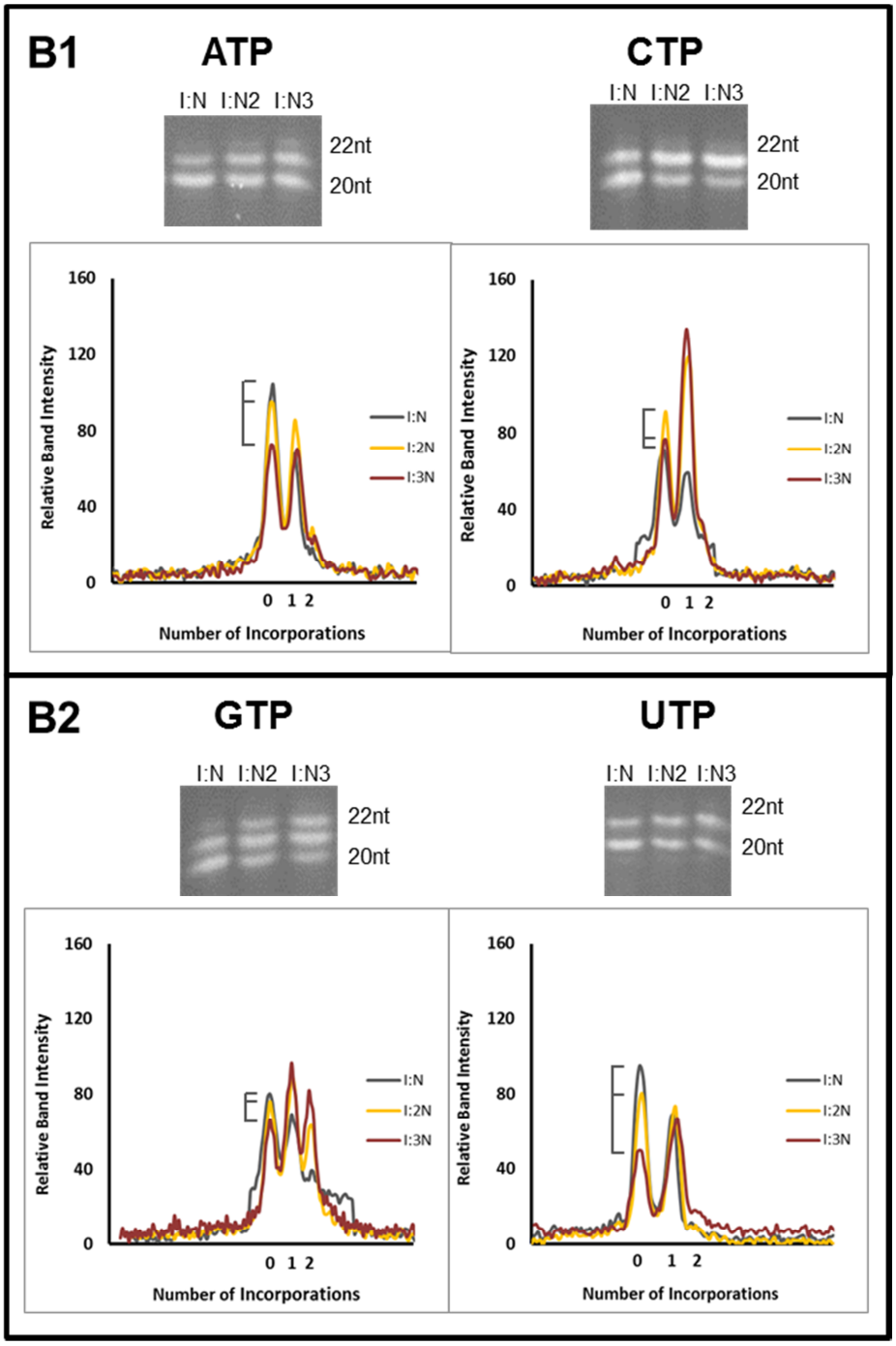

Fig. 12.2 – TdT polymerization reaction with increasing concentration of NTPs. 12% denaturing polyacrylamide gel image of 1:1, 1:2 and 1:3 incorporation of: ATP and CTP (B1), GTP and UTP (B2); gel band density profile for each nucleotide, where individual peaks from left to right represent different product sizes (20-22 nucleotides); the marker near the 20 nucleotide peak illustrates the consumption of initiator molecules with increasing concentration of NTPs.

As a result, because earlier reports stated that TdT does not discriminate between deoxynucleotides and ribonucleotides,[36] the next step was the verification of how did the reaction proceed with increasing concentrations of NTPs. The experimental plan was to substitute dNTPs for NTPs in the polymerization reaction and evaluate the enzyme's behavior. Only then it would be possible to compare the experimental results with the ones obtained with caged ATP in terms of polymerization efficiency.

These results, obtained as indicated previously, are displayed above in figure 12.2. In this case, it is observed a maximum of 2 incorporations in all four experiments. As observed in figure 12.1, the incorporation profiles of ATP and GTP display opposite results. Whilst ATP presents the lowest initiator first incorporation, GTP is the only one that presents a clear band representing the second nucleotide incorporation into the initiator chain. Moreover, CTP and UTP present similar results with a mean value of incorporation of 1 nucleotide per initiator molecule. By comparing figure 12.1 with figure 12.2, one notices that this enzyme does not easily accommodate nucleic acids terminated with ribonucleotides as the average result of figure 12.2 was a single incorporation of a ribonucleotide. Hence, in studies where TdT is forced to incorporate ribonucleotides, we still use a deoxyoligonucleotide initiator. Interestingly, when NTPs were used, the results showed a significant decrease in the efficiency of the reaction while we still could group them into the same categories. This is consistent with previous results where a 2 to 9-fold preference from TdT to the incorporation of dNTPs is reported.[13]

All these combined together validate once more the theory of a natural structural selectivity of TdT towards different nucleotides present in the system. In fact, besides TdT, other DNA polymerases present a specific mechanism of nucleotide selection that is fundamental for their biological role.[5] Such differences in incorporation profiles have to be considered when modelling the kinetics of this reaction.

3.2 Light activation of *in vitro* polymerization with TdT

Caging compounds are biomolecules that are unable to participate into chemical reactions unless an external stimulus is performed in order to cleave the photoremovable protecting group. Due to its properties, caging compounds for biological molecules have been used for decades in order to achieve a stringent control of specific reactions.[25] In this particular case such level of control can be attained using a caged molecule that, when exposed to light irradiation, is able to release a key ingredient in polimerization reactions – figure 13. Without its release, reaction does not occur. In this project derivatives of 4-methyl-coumarines will be used in the form of DEACM-ATP to achieve such level of control. Upon conjugation with a nucleotide (in this case, ATP) DEACM-ATP is formed, rendering this nucleotide biologically inert. DEACM presents a high extinction coefficient near 390 nm, and the phosphodiester bond between the two compounds is prone to be photocleaved.[32] So, when the caged molecule is exposed to light irradiation at this wavelength the phosphodiester bond breaks and the nucleotide is released to be used as a substrate for the polymerase. Although theoretically this concept seems simple, in practice the decaging system needed to be tested before using these molecules in polymerization reactions.

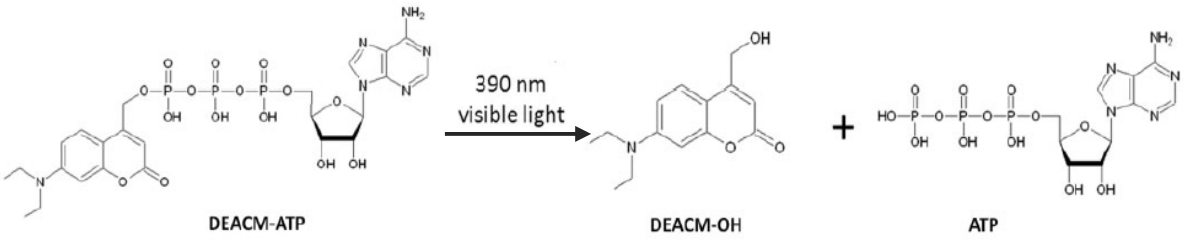

Fig. 13 – Light activated hydrolysis of DEACM-ATP and release of ATP molecules. Upon a light stimulus of 390 nm the ester bond between DEACM and the gamma phosphate group of the ATP molecule is disrupted. ATP is released and DEACM-OH is generated. Adapted from Pinheiro *et al.* 2008 [32].

3.2.1 Monitorization of Light induced decaging of ATP

Upon 390nm irradiation of a DEACM-ATP solution, the bond between DEACM and ATP molecules is cleaved and free ATP and DEACM-OH are generated[32] (Fig. 13) – this

reaction is referred here as the decaging of ATP. In order to achieve a maximum efficiency in the photolysis of DEACM-ATP, the analysis of this solution upon different irradiation times was performed. Standard solutions of DEACM-ATP before irradiation (100 µM) and DEACM-OH (100 µM) were prepared to calibrate the HPLC analytical method created to survey the release of ATP. The identification of both species is accomplished by the retention time observed when the standard solutions are injected into the HPLC separately, as shown in figure 14.

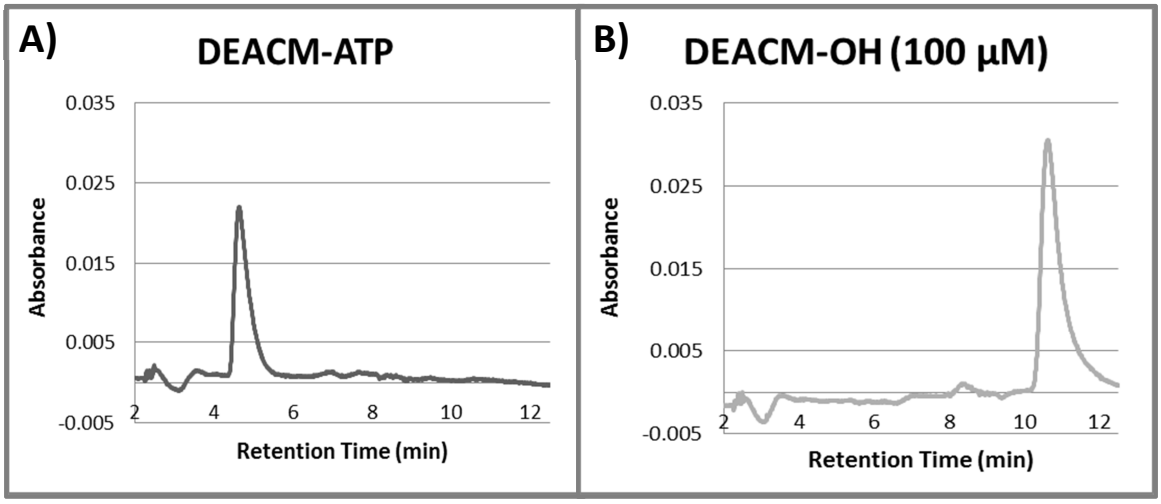

Fig. 14 – HPLC chromatograms of DEACM-ATP (A) and DEACM-OH (B) standard solutions (100 µM). The retention time at which a solution appears when analysed individually defines the pattern that identify that solution when a mixture of compounds is analysed.

A gradient elution method was performed to achieve a maximum separation of both compounds and the results obtained correspond to the displayed chromatograms. DEACM-ATP molecules are retained less time in the column and reach the detector within 4.5 minutes after the beginning of the analysis when the elution gradient is stabilizing in 90% of methanol (Fig. 14A). The ATP molecule itself is undetected by the selected method as it presents no retention in a polymeric reversed-phase HPLC column (PLRP-S). However, DEACM-OH, a secondary product of the reaction, is bound by the column with more affinity than DEACM-ATP and presents a retention time at 10.5 minutes. As a product from light induced decaging, this molecule represents a perfect standard for this assay (Fig. 14B). The elevated selectivity verified in the chromatography assay accounts for observed

differences in the retention times of both molecules, allowing for a good separation of these compounds.

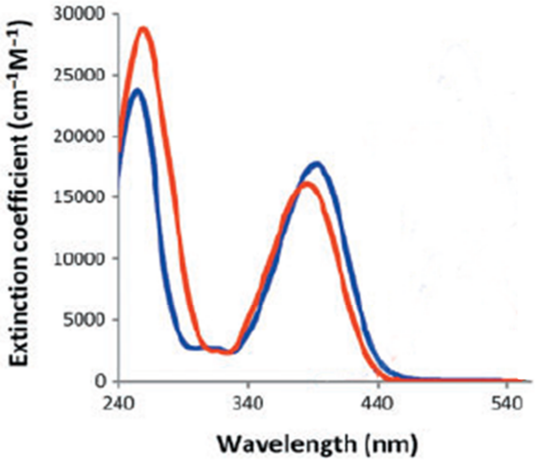

Fig. 15 – Absorption spectra of DEACM-ATP (blue) and DEACM-OH (red). Adapted from Pinheiro *et al.* 2008 [32].

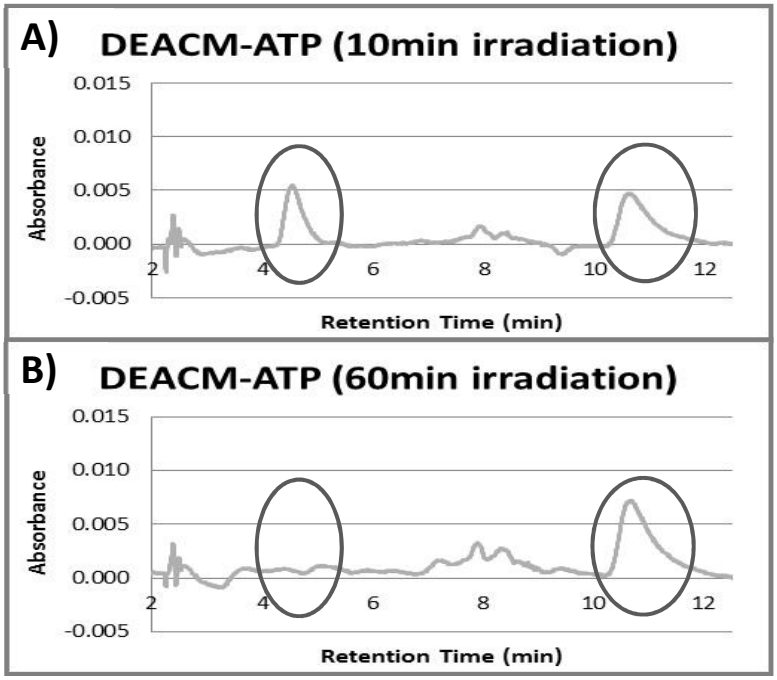

Fig. 16 – HPLC chromatograms of DEACM-ATP solutions after 10 minutes (A) and 60 minutes (B) of irradiation. The left circle highlights DEACM-ATP (4.5 minutes) while the right circle represents the DEACM-OH elution peak (10.5 minutes). Only after 60 minutes of irradiation is observed a complete hydrolysis of a 100 µM solution of DEACM-ATP.

When subjected to light irradiation, the concentration of DEACM-ATP is an important factor to consider because higher periods of irradiation time are needed for more concentrated solutions in order to maintain a similar yield of nucleotide release from the coumarin. This can be explained because both DEACM-ATP and DEACM-OH absorb energy in the 390 nm range of the spectrum and are competing for each light photon (Fig. 15). Therefore, it is not unusual that after 10 minutes of irradiation, with an I_0 of 4.97×10^{-9} Einstein/cm^3/s, there is still some DEACM-ATP left in solution, which was not photolised (Fig. 16A). Only after 60 minutes of irradiation all the DEACM-ATP present in solution was converted into free ATP and DEACM-OH (Fig. 16B). Consequently, in subsequent studies that require the presence of free ATP in solution, a 60 minutes light irradiation of a 100 µM DEACM-ATP solution is performed.

3.2.2 Inhibition of TdT polimerization by DEACM-OH

As mentioned beforehand, when light mediates the DEACM-ATP hydrolysis, a secondary product is formed: DEACM-OH. Previous data reports that there is a chance this compound inhibits polymerization in the same manner as it interferes with in vitro transcription.[32] However, the mechanism by which this inhibition effect is achieved is still unknown. A putative enzyme inhibition mediated by DEACM-OH alters the approach reliability, and so an inhibition assessment was carried out. The inhibition study was performed as described in section 2.3 using ATP as a free nucleotide source. Several concentrations of DEACM-OH were used corresponding to different I:N ratios and the results were obtained following the same experimental design as in previous sections. In figure 17 we have represented two results obtained from this experiment. In each experiment 25 pmol of initiator molecule were used. As observed in previous studies, with ATP molecules, TdT easily incorporates one nucleotide into the initiator sequence and a second incorporation starts to appear in slight quantities. When 50 pmol of DEACM-OH (I:N - 1:2) is present, there is no visible inhibition of the polymerization reaction confirmed by the overlay of the reaction profiles with and without DEACM-OH. However, when 150 pmol of DEACM-OH (I:N - 1:6) is added to the polymerization reaction, there is a small

decrease in the efficiency of the reaction, observed with the decrease in the relative intensity of the correspondent 21 nucleotide band and the disappearance of the small 22 nucleotide band. This clearly indicates that 150 µM of DEACM-OH interferes with the polymerization reaction of TdT.

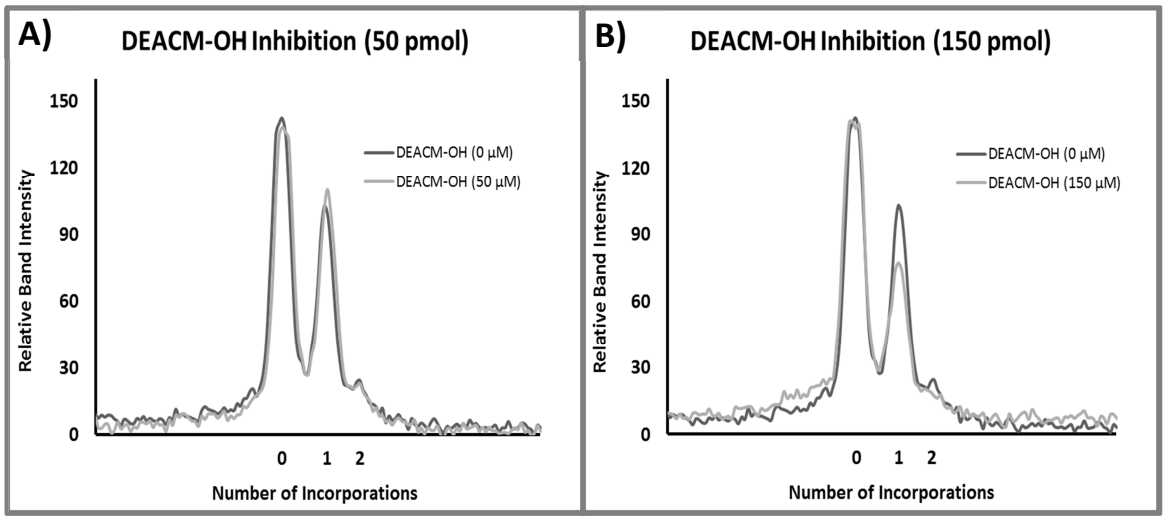

Fig. 17 – Gel band profile for TdT polymerization reactions with ATP as a nucleotide source and DEACM-OH at 50 µM (A) and 150 µM (B). The polymerization reaction is not affected by the presence of 50 µM of DEACM-OH. However, with 150 µM of DEACM-OH it is observed a decrease in the reaction yield.

3.2.3 TdT's Polymerization reaction with DEACM-ATP

Non-irradiation studies with DEACM-ATP and TdT were performed in order to assess (in the dark) the integrity of the caging molecule as well as the efficiency of the caging. Therefore, increasing concentrations of DEACM-ATP were added to the enzymatic reaction where no nucleotide source was present.

Figure 18 shows the effect of DEACM-ATP on the referred polymerization reaction. Up to a concentration of 200 pmol, not only there was no visible sign of nucleotide incorporation, but also the differences in pixel intensity of all samples are not significant. However, we still wanted to know if increasing the concentration to 250 pmol (1:100) of DEACM-ATP would produce a similar result. Curiously, there was a small observed incorporation which

leads to the belief that only after 250 pmol there is enough amount of ATP to yield an incorporation reaction without DEACM-ATP irradiation (Fig. 19). We then concluded that a leakage phenomenon is not present and only with higher quantities of DEACM-ATP (above the range of the present work) a minor quantity of free ATP is present in solution, resulting in the formation of small amounts of 21 nucleotide products. Therefore, we found the caging system to be effective and suitable to be used in subsequent experiments to control polymerization through light.

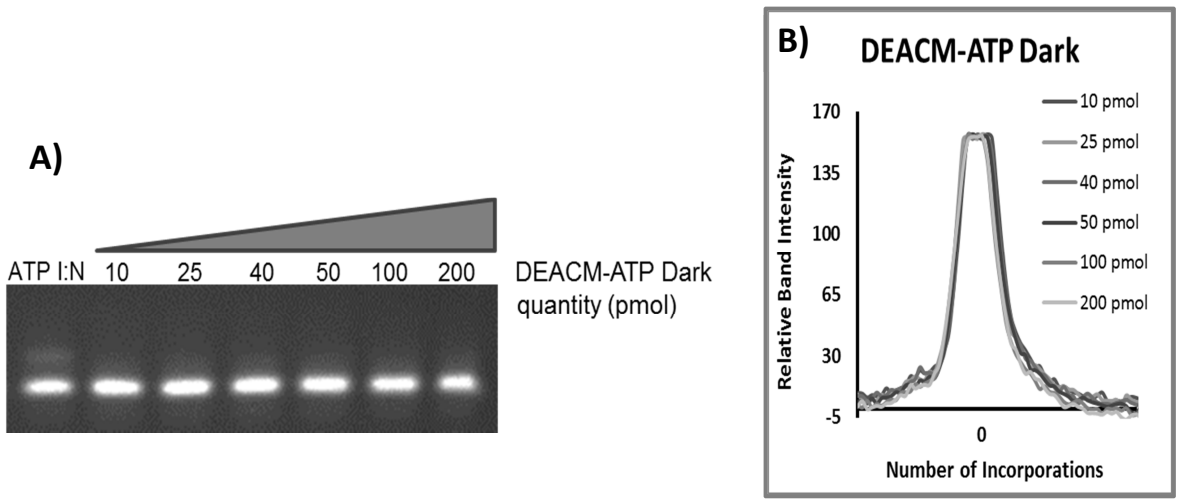

Fig. 18 – Effect of DEACM-ATP into the polymerization reaction. Increasing the concentration of DEACM-ATP to 200 µM does not influence the nucleotide incorporation reaction. Results presented in a 12% polyacrylamide gel (A) and in a plot of gel density approach (B).

We have successfully demonstrated that TdT incorporates both deoxynucleotides and ribonucleotides with highest affinity for deoxynucleotides; DEACM-ATP (100 µM) is entirely cleaved after 60 minutes under 390 nm irradiation with an I_0 of 4.97 x 10^{-9} Einstein/cm^3/s; before irradiation, DEACM-ATP does not participate into TdT's enzymatic reaction when no nucleotide source is present; and DEACM-OH only affects the polymerization efficiency when is above 150 pmol. The next step is the combination of all this factors in attempt to achieve an integrative perspective related to this subject.

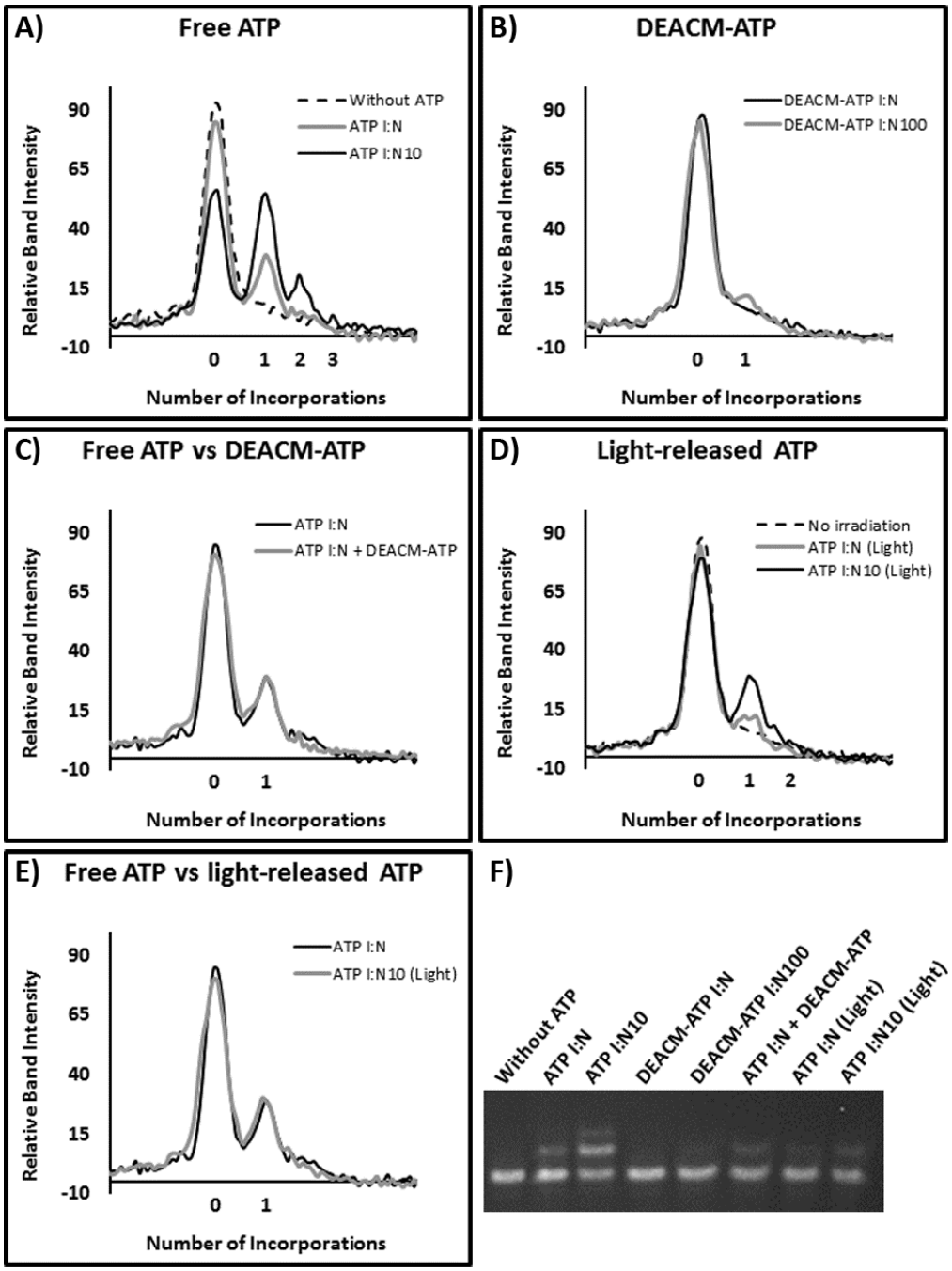

Fig. 19 – TdT's Polymerization reaction with DEACM-ATP irradiation. Incorporation reaction controls with free ATP (A), and without irradiation of DEACM-ATP (B) are presented. A comparison between incorporation of free ATP with and without the presence of DEACM-ATP was performed as an experience control (C). Incorporation reaction with DEACM-ATP as a nucleotide source is presented (D); with increasing concentration of DEACM-ATP longer products are formed. Overlaping of the profile of incorporation from free ATP incorporation (1:1) and light activated ATP incorporation (1:10) was performed (E). Data obtained from analysis of a 12% polyacrylamide gel where the result from the polymerization reaction of TdT is displayed (F).

Results from incorporation of free ATP and light activated ATP with different I:N ratios (1:1 and 1:10) and related reaction controls were obtained and compared – Figure 19. First, it is important to mention that DEACM-ATP does not interfere in the polymerization reaction when free ATP is also present, as we observe in figure 19C. Although there is a visible incorporation of ATP after DEACM-ATP is irradiated for 60 minutes at 390 nm, considering that DEACM-ATP is totally cleaved, it is observed in the above figure that the incorporation reaction is less efficient when compared to the incorporation reaction with free ATP. Considering the DEACM-OH inhibition observations, it is easy to justify the slow incorporation of ATP in I:N10 case because of a light absorption interference from DEACM-OH. However, a lower incorporation of ATP is not as expectable when an equimolar concentration of initiator and irradiated DEACM-ATP is used because smaller amounts of DEACM-OH are present in solution.

Interesting is the observation that the incorporation pattern of light activated ATP matches the pattern of free ATP incorporation when this molecule is 10 times more concentrated than the initiator itself – Figure 19E. Considering these results, we have to reflect on the efficiency of the decaging reaction of ATP. If a complete photolysis reaction is achieved after 60 minutes of irradiation and the concentration of the irradiated solution is calculated assuming that one molecule of DEACM-ATP originates one free ATP, the results cannot be explained, since DEACM-OH does not affect the reaction at these concentrations and the HPLC chromatogram shows that all DEACM-ATP reacted after 60 minutes. These results clearly show that only 10% of the reacted DEACM-ATP yields free ATP, after all DEACM-ATP is photolysed. This is the first evidence that DEACM-ATP has an additional photochemical pathway yielding a photoproduct, not identified by HPLC, which accounts for 90% of the irradiated DEACM-ATP.

3.3 Prediction of the system's evolution: a Kinetic Model based on the Chemical Master Equation

The mathematic model used to predict the system's evolution is based on the statistical probability of an enzymatic reaction to occur in a specific time interval. Such probability of occurrence depends directly on: the nucleotide quantity, approximately described in number of molecules (to define the number of possible incorporations); the number of available enzymes; and an arbitrary kinetic constant (that defines the velocity by which the nucleotides are incorporated). All of these parameters taken together define the product formation rate. Here is presented and discussed the methodology that led to the comparison between the biological results presented above and the model outputs. All of the program detailed aspects will not be discussed here as they do not integrate this work's purpose.

A mathematic model that considers the characteristics presented above is a necessary tool to comprehend the evolution of the polydispersion that reflects the incorporation profile of the DNA polymerase TdT (the main focus of this study). Direct observation of all the previously discussed results show us that such polydispersion is only relevant when this enzyme incorporates dNTPs into a deoxyoligonucleotide initiator (Fig. 12.1), as it exhibits a very limited number of incorporations when using NTPs as a substrate (Fig. 12.2). Therefore, the following discussion is only relevant when we consider the results for dNTPs. Moreover, the mathematical model created for this project does not accommodate discrimination between dNTPs and NTPs in terms of enzymatic affinity.

3.3.1 Experimental results vs Simulation Model

In order to assist a comparison between the experimental data and the stochastic model simulations, data from peak area quantification had to be converted into incorporation statistics data. Considering the sum of the peak area values, each incorporation peak was converted into percentage of incorporation. Then, two statistical distributions were used to be fitted into the experimental data: Binomial and Poisson. Because each incorporation reaction is performed by independent enzymes at the same time, we consider all

incorporation reactions as independent events that may, or may not result in the successful incorporation of each nucleotide. The probability of incorporation of a certain number of nucleotides is given by the Binomial statistic distribution, defined by *p*. On the other hand, the Poisson statistic distribution indicates the average number of nucleotides incorporated by TdT, defined by λ, at the end of the reaction, considering the incorporation probability of each incorporated nucleotides, present in 1, 2, 3, etc. incorporation sites, given by the binomial distribution. The application of this statistical approach is an important factor when a comparison between experimental results and model simulations is in order.

To assess if the model could accurately predict the progression of the enzymatic reaction, a number of simulations was performed. Several k values (the arbitrary constant related to the enzymatic velocity) were tested with different Initiator/Nucleotide ratios to replicate the experimental conditions tested in the lab. With these initial conditions the program calculates the propensity function $aj(x)dt$ that depends on the mean value of the *poisson* probability distribution. Then, several values of probability of incorporation are generated and a column plot of number of molecules versus product size appears. In this plot we can also evaluate the number of initiator and nucleotide molecules that were theoretically consumed in the course of the simulation reaction. With the obtained results from the simulation, a percentage of nucleotide incorporation was calculated as performed with the experimental results. Then, a plot of incorporation percentage versus I:N was created to fit the model previsions into the biologic data. To complete each analysis, several simulations were performed where different I:N ratios were tested – Fig. 20. The comparison itself was not only performed in order to identify the k value that best suited the observed results, but also, to identify the mean value of the poisson distribution that corresponded to the average nucleotide incorporation value of the experimental conditions.

3.3.2 Polydispersion polymerization study

With the intention to evaluate the accuracy of the mathematic model regarding a prediction of TdT's progression during the enzymatic reaction, a comparison between the probability of incorporation obtained experimentally and the simulation results was performed.

Moreover each *k* value applied in this experiment represents an indirect measure of the incorporation velocity of TdT in the presence of the four nucleotides. This arbitrary *k* value results from multiplying the nucleotide concentration for the obtained incorporation probability. Therefore, a graph was created for each dNTP assay, where the *p* value was plotted against the I:N ratio – Figure 20.

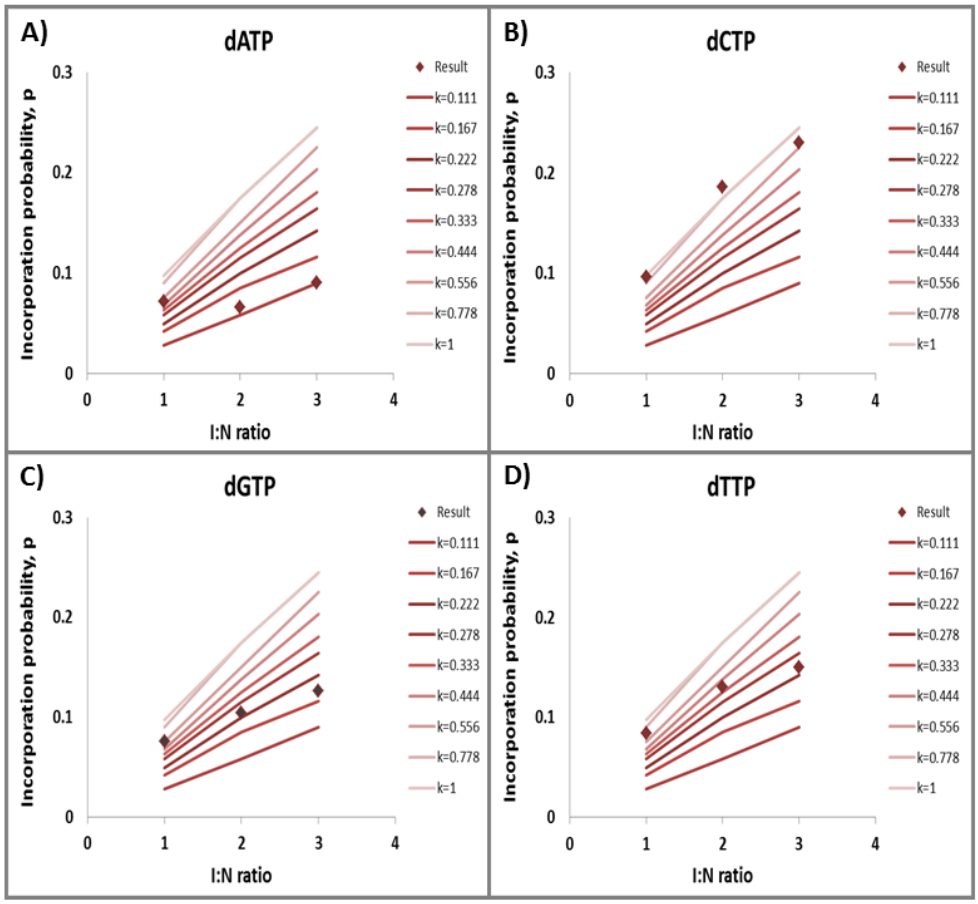

Fig. 20 – Fitting of the proposed biological system of dNTPs incorporation with the created mathematic model. The dots represent the experimental results and the lines represent a statistical simulation based on different k values for each dNTP; as an arbitrary constant the k values are indicated in function of the higher value.

When we consider the results presented in figure 20 we can observe that the profile of incorporation when dCTP is used as a nucleotide source (Fig. 20B) accurately describes a stochastic reaction characterized by only one enzymatic constant. A stochastic profile of incorporation evolves by an increase of the incorporation probability when more substrates

are available to be incorporated, in this case nucleotides. Observing the results obtained with dGTP and dTTP (Figs. 20C and 20D) there is a stabilization of the enzymatic velocity for increasing concentration of nucleotides. This effect is consistent with an enzymatic reaction that has almost achieved its maximum product formation rate. However, the case of dATP is different from the others (Fig. 20A). Its incorporation probability is stabilized from the beginning, and continues so with increasing concentrations of substrate. This observation is not only different from a stochastic incorporation profile, but also justifies the low incorporation rate of this nucleotide. In order to efficiently use the model to predict the incorporation progression such differences amongst dNTPs' incorporation rate have to be considered.

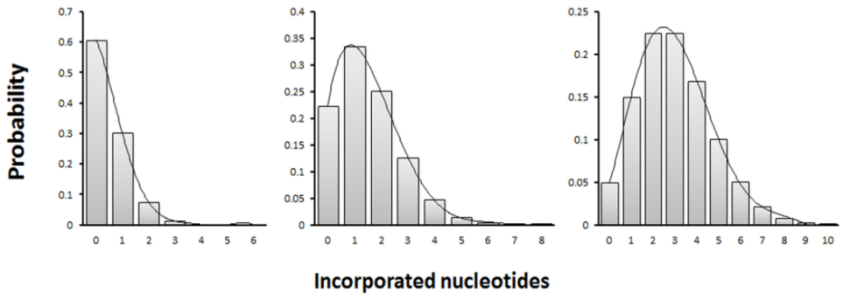

Fig. 21 – Evolution of the Poisson distribution with the increase in the mean value of nucleotide incorporation. The values used do not represent any obtained result; they only serve the purpose of evaluation of the evolution of this probability distribution.

As we mentioned early, when looking to the experimental data, we also have also to consider the evolution of product size distribution with increasing concentrations of the limiting reagent, i.e. nucleotides. This defines the profile of incorporation of each nucleotide. When we look back to the dNTP incorporation results we observe that the profiles of incorporation are all different (Fig. 12.1). Once more, the profile that most resembles the progression of a poissonian distribution is the profile of dCTP where there is a clear change in the mean value of product sizes with increasing concentrations of nucleotide in solution as demonstrated in figure 21. Such profile is the typical one for a stochastic distribution. The profile of incorporation of dTTP and dGTP is considered to resemble a poisson distribution with low probability of incorporation where 20, 21 and 22 nucleotide-long products are predominant. However, dATP presents a different profile

from the other nucleotides. It is observed that with increasing concentration of nucleotides the incorporation profile remains the same and no more than three nucleotides are incorporated in each initiator.

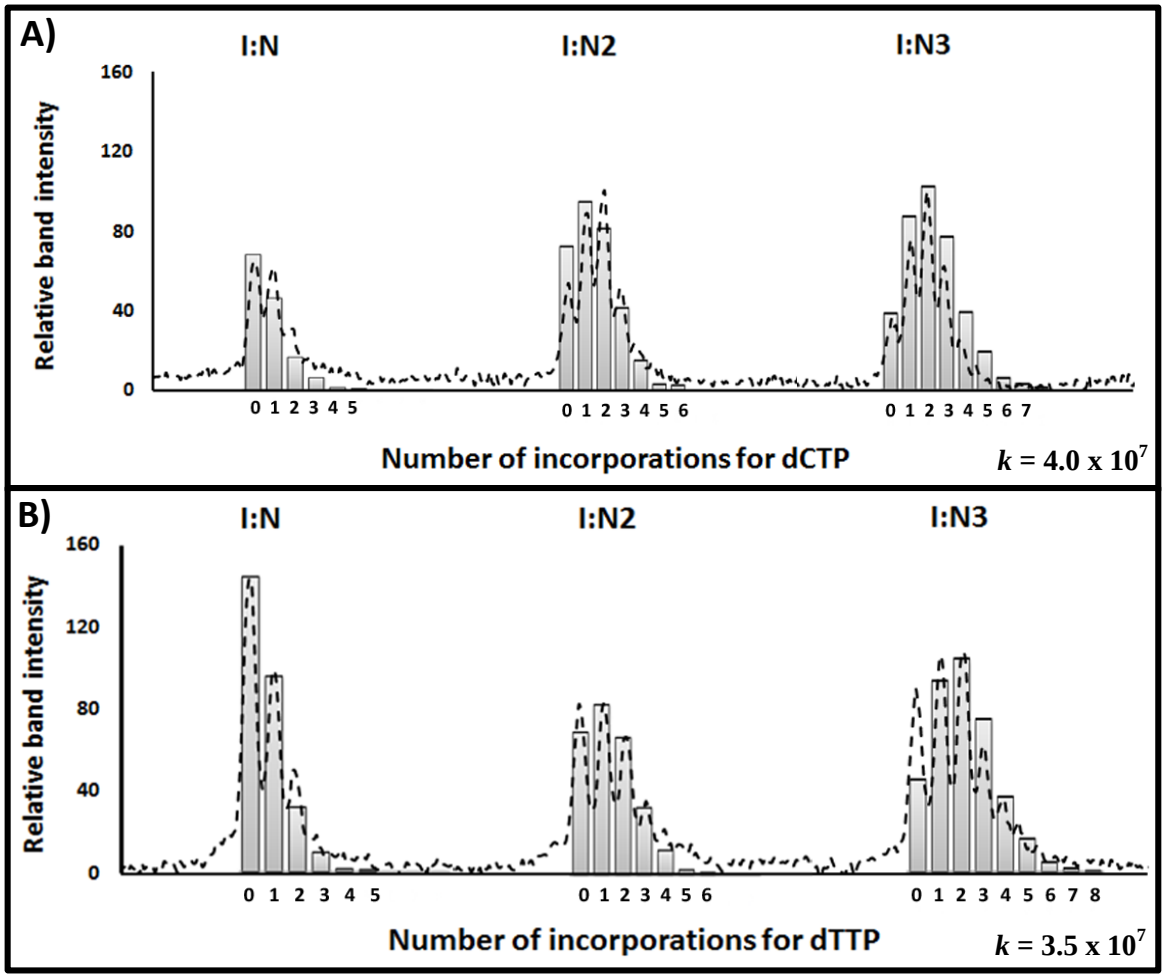

Fig. 22.1 – Comparison between experimental data and mathematical model output data for dCTP and dTTP. Probability of incorporation predicted using the mathematical model merged with relative band intensity profiles obtained using the pixel quantification technique.

In figure 22.1 is represented a comparison between the profiles of incorporation of each dNTP and the profile given by the simulation model that best fitted the experimental results. Each simulation was obtained with different values of *k*, the arbitrary velocity constant. The same stoichiometric conditions used in the experimental approach were introduced in the simulation model. Considering all the results presented in figure 22.1,

dCTP incorporation profile is the one that the adopted simulation model best describes and the one which suits the highest constant value. The incorporation profile for dTTP, on the other hand, despite its resemblance with a stochastic incorporation pattern, is suited by a lower arbitrary kinetic constant. The other two cases presented in figure 22.2, namely dGTP and dATP, were considered not to follow a stochastic kinetic profile through direct observation of their incorporation profiles. Therefore, the fittings were made with a chosen *k* value of 3.5 x 10^7 in order to observe a progressive deviation from the simulation model to the experimental results. The behavior of TdT in the presence of dATP will now be discussed in terms of the Michaelis-Menten model of enzyme kinetics.

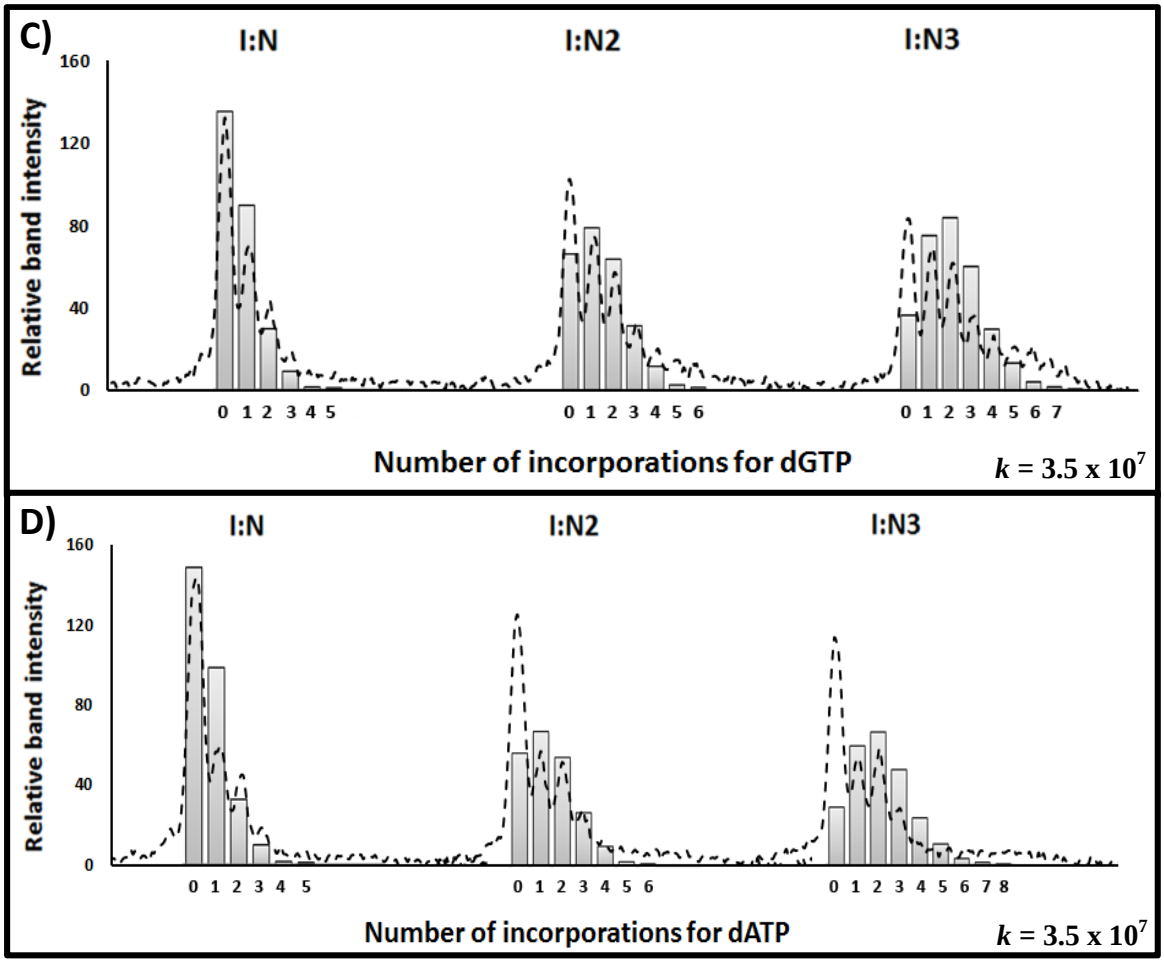

Fig. 22.2 – Comparison between experimental data and mathematical model output data for dGTP and dATP. Probability of incorporation predicted using the mathematical model merged with relative band intensity profiles obtained using the pixel quantification technique.

Regarding the classic Michaelis-Menten equation we can observe that the velocity of each enzymatic reaction is dependent on the substrate concentration.[22,23] In the case of dATP we do not observe such correlation. This might be explained if we consider that the maximum velocity of the reaction was rapidly reached (higher initial velocity) for small dATP concentrations. In other words, we might consider that the velocity of the incorporation reaction of dATP has already reached a constant value. Then, the availability of more substrate molecules would not interfere with the reaction course as it is somehow limited to a constant rate of polymerization. Such independence towards the substrate concentration is also observed when the reaction cannot progress efficiently to the next cycle of incorporation.

Two hypotheses might be considered that could justify the absence of higher size products when there is free dATP and initiator available. The first one is related to the ability of TdT to efficiently release a product molecule that resulted from a deoxynucleotide insertion in the 3'-end of an initiator molecule - dissociation of the enzyme-product complex. The obtained results suggest that this dissociation step does occur with some nucleotides as we observe multiple incorporations per initiator sequence with dCTP and dGTP, for example, but the rate of this occurrence might be very slow in the case of dATP. In this scenario not only the enzyme is unavailable to proceed to the next cycle of incorporations, but also the progression of the reaction reaches a physical state when every enzyme-product complex is formed but is not able to dissociate, thus, no more usable enzyme molecules are available. So, in this case we might conclude that the product formation is dependent on the enzymatic concentration. The other hypothesis that is proposed is related to the ability of TdT to efficiently recognize and bind an initiator chain that was previously polymerized with a specific dNTP. In other words, the enzyme may be facing an inability of association with the substrate with a consequent hindrance in the formation of the enzyme-substrate complex. However, in order to validate this assumption, we would need to consider dATP as a particular case, as none of the polymerization reactions using other dNTPs seem to exhibit the same deficiencies. Moreover, published literature states that TdT, as well as other DNA polymerases, present a high affinity for dNTPs. It was previously described a

discrimination against the utilization of dATP by TdT[11] however, the cause for this discrimination is still unknown.

To efficiently adapt a mathematic model to the biological system two things have to be considered. First, the model has to contemplate each one of the nucleotide incorporations as an independent event. Secondly, all the enzymatic characteristics that result in a substrate selectivity have to be bypassed in order to achieve a constant enzymatic velocity that described the entire reaction. One example is the “lariat-like” loop that accounts for TdT’s nucleotide specificity. Another example is the adopted mixture of co-factor molecules used to assist the nucleotide incorporation by TdT. Moreover, a single-molecule kinetic approach related to a pulse irradiation method would be considered.

4. Conclusions and Future perspectives

The main challenge presented in this project was the development of an efficient system for oligonucleotide synthesis directly in solution, recurring to the template-independent DNA polymerase Terminal Deoxynucleotidyl Transferase. The concept relies on the use of monochromatic light to control the release of nucleotides in solution. 4-methylcoumarin derivatives are used as nucleotide cages, blocking the enzyme-substrate recognition, which becomes available upon a light stimulus of a determined wavelength.

Without control of the reaction, it is not possible to ensure a specific product length with a desired sequence, because a mixture of enzymes is present in solution, each possessing its own polymerization rate constant. Therefore, it is important to analyze the polymerization reaction from a single-molecule point of view, requiring stochastic kinetic methods to evaluate the state evolution of the system. Computational models have been regarded as the most appropriate tool to achieve such level of accuracy. Moreover, the Chemical Master Equation can be used to describe the progression of enzymatic polymerization reactions. Regarding this insight, it becomes possible to propose a kinetic mechanism and to achieve the optimal conditions for a controlled enzymatic incorporation of nucleotides into a short pre-existing oligonucleotide chain. This was the main goal of the present work.

Different stoichiometric conditions were tested for incorporation of dNTPs and NTPs to acquire enough data to create an accurate kinetic model. These experiments confirmed that TdT does not discriminate between nucleotides, but preferentially incorporates dNTP into the 3'-end of an oligonucleotide initiator. The polydispersion that was the central study of this project was only observed when dNTPs were used as building blocks for polymerization. On the other hand, it was observed that this enzyme is an inefficient polymeric agent when a ribonucleotide is in the 3'-end position as we witness no more than 2 incorporations when NTPs are used as building blocks. Therefore, a DNA-RNA hybrid was synthesized with TdT to evaluate the enzymatic progression with NTPs. At this point, ATP caged with DEACM was introduced to the system, towards the study of light control of polymerization.

Several reaction controls were made to test the coumarin molecule's effectiveness as a caging agent. It was concluded that DEACM was an efficient caging agent (non-irradiated DEACM-ATP did not interfere with the polymerization reaction when its quantity is below 250 pmol). Moreover, with an I_0 of 4.97 x 10^{-9} Einstein/cm^3/s, after 60 minutes of irradiation all DEACM-ATP were converted into free ATP and DEACM-OH. However, inhibition was observed when DEACM-OH, a secondary product of the DEACM-ATP hydrolysis, was present in the reaction at 150 pmol. The inhibition concentration corresponded to a 1:6 Initiator:Nucleotide ratio. This ratio would result in the incorporation of only 6 nucleotides, considering each initiator being polymerised at the same processivity rate. Regarding the main purpose, an incorporation of only 6 nucleotides per initiator is unsuitable and therefore a mechanism to prevent such inhibitory effect must be considered in future experiments.

The results obtained from the densitometry approach related to the incorporation of dNTPs were converted into statistic data and compared to the output obtained from the mathematic model. It was observed that the profile of incorporation of dTTP and dCTP were the most easy to predict because they followed a stochastic kinetic profile with increasing concentration of nucleotides. On the contrary, the profile of dATP and dGTP was not related to the one obtained with the model. Moreover, the rate of incorporation of dATP is the lowest of the four deoxynucleotides. When comparing the arbitrary velocity rate constant for each nucleotide incorporation it is observed a variation within each incorporation site, therefore difficulting a general prediction of incorporations. A valid hypothesis is the assumption from which the model originated, stating that the incorporation velocity does not change in the course of polymerization. This demonstrates that not only we have to think in a way to overcome the selectivity that TdT presents to each nucleotide, but also the differences among the velocities of incorporation nucleotides in each position. It is already known that the "lariat-like" loop characteristic from TdT and the selected co-factors are responsible for the nucleotide selectivity, but they may also be responsible for the observed velocity variations. However, more studies are needed to confirm this assumption. An approach based on enzymatic mutagenesis has been considered as a valid means to overcome the encountered obstacles in this project. This work allowed one more step into the understanding of the TdT's system of incorporation

towards the development of an efficient enzymatic system for synthesizing short oligonucleotide chains.

Overall, this project allowed the creation of an efficient system for analysis of enzymatic polymerization reactions using polyacrylamide gel electrophoresis with subsequent gel densitometry approaches, statistical distribution analysis and mathematic simulation models in order to understand and predict *in vitro* enzymatic synthesis of nucleic acids.

5. References

1. Steitz, T. A., DNA Polymerases: Structural Diversity and Common Mechanisms. *The Journal of Biological Chemistry* **1999**, *274* (25), 17395–17398.

2. Bartlett, J. M.; Stirling, D., A Short History of the Polymerase Chain Reaction. *Methods in Molecular Biology* **2003**, *226*, 3-6.

3. Sanger, F.; Coulson, A. R., A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **1975**, *94* (3), 441-448.

4. Losick, R., In Vitro Transcription. *Annual Review of Biochemistry* **1972**, *41*, 409-446.

5. Hingorani, M. M.; O'Donnell M., DNA Polymerase Structure and Mechanisms of Action. *Current Organic Chemistry* **2000,** *4* (9), 887-913.

6. Itakura, K.; Rossi J. J.; Wallace R. B., Synthesis and Use of Synthetic Oligonucleotides. *Annual Review of Biochemistry* **1984,** *53*, 323-356.

7. Krayevsky, A. A.; Victorova L. S.; Arzumanov A. A.; Jasko M. V., Terminal deoxynucleotidyl transferase: catalysis of DNA (oligodeoxynucleotide) phosphorylation. *Pharmacology & Therapeutics* **2000,** *85*, 165–173.

8. Ramadan, K.; Shevelev, I. V.; Maga, G.; Hübscher, U., De Novo DNA Synthesis by Human DNA Polymerase λ, DNA Polymerase μ and Terminal Deoxyribonucleotidyl Transferase. *Journal of Molecular Biology* **2004,** *339*, 395–404.

9. Deiters, A., Principles and Applications of the Photochemical Control of Cellular Processes. *ChemBioChem* **2010,** *11*, 47-53.

10. Chang, L. M. S.; Bollum, F. J., Enzymatic Synthesis of Oligodeoxynucleotides. *Biochemistry* **1971,** *10*, 536-542.

11. Motea, E. A.; Berdis, A. J., Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. *Biochimica et Biophysica Acta* **2010,** *1804*, 1151–1166.

12. Wuite, G. J. L.; Smith, S. B.; Young, M.; Keller, D.; Bustamante, C., Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* **2000,** *404*, 103-106.

13. Fowler, J. D.; Suo, Z., Biochemical, Structural, and Physiological Characterization of Terminal Deoxynucleotidyl Transferase. *Chemical Reviews* **2006,** *106*, 2092-2110.

14. Johnson, K. A., Conformational Coupling in DNA Polymerase Fidelity. *Annual Review of Biochemistry* **1993**, *62*, 685-713.

15. Delarue, M.; Boulé, J. B.; Lescar, J.; Expert-Bezançon, N.; Jourdan, N.; Sukumar, N.; Rougeon, F.; Papanicolaou C., Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *The EMBO Journal* **2002,** *21* (3), 427-439.

16. Guerra, C. E., Analysis of oligonucleotide microarrays by 3′ end labeling using fluorescent nucleotides and terminal transferase. *BioTechniques* **2006,** *41*, 53-56.

17. Di Santo, R.; Maga, G., Human Terminal Deoxynucleotidyl Transferases as Novel Targets for Anticancer Chemotherapy. *Current Medicinal Chemistry* **2006,** *13*, 2353-2368.

18. Futami, K.; Furuichi, Y., Quantitation of full-size small interfering RNA by tailing with terminal deoxynucleotidyl transferase and reverse transcription-polymerase chain reaction analysis. *Analytical Biochemistry* **2009,** *385*, 386–388.

19. Joyce, C. M.; Steitz, T. A., Function and structure relationships in DNA Polymerases. *Annual Review of Biochemistry* **1994,** *63*, 777-822.

20. Brown, J. A.; Suo, J., Unlocking the Sugar ‘Steric Gate’ of DNA Polymerases. *Biochemistry* **2011,** *50* (7), 1135–1142.

21. Kou, S. C.; Cherayil, B. J.; Min, W.; English, B. P.; Xie X. S., Single-Molecule Michaelis-Menten Equations. *The Journal of Physical Chemistry* **2005,** *109*, 19068-19081.

22. English, B. P.; Min, W.; van Oijen, A. M.; Lee, K. T.; Luo, G.; Sun, H.; Cherayil, B. J.; Kou, S. C.; Xie X. S., Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nature Chemical Biology* **2006,** *2* (2), 87-94.

23. Qian, H.; Elson, E. L., Single-molecule enzymology: stochastic Michaelis–Menten kinetics. *Biophysical Chemistry* **2002,** *101–102*, 565–576.

24. Gillespie, D. T., Stochastic Chemical Kinetics. *Handbook of Materials Modeling* **2005,** 1735–1752.

25. Lee, H.-M.; Larson, D. R.; Lawrence, D. S., Illuminating the Chemistry of Life: Design, Synthesis, and Applications of "Caged" and Related Photoresponsive Compounds. ACS Chemical Biology **2009,** *4* (6), 409-427.

26. Young, D. D.; Deiters, A., Photochemical control of biological processes. *Organic & Biomolecular Chemistry* **2007,** *5*, 999-1005.

27. Kaplan, J. H.; Forbush, B.; Hoffman, J. F., Rapid Photolytic Release of Adenosine 5'-Triphosphate from a Protected Analogue: Utilization by the Na:K Pump of Human Red Blood Cell Ghosts. *Biochemistry* **1978,** *17*, 1929-1935.

28. Deiters, A., Principles and Applications of the Photochemical Control of Cellular Processes. *ChemBioChem* **2010,** *11*, 47-53.

29. Krock, L.; Heckel, A., Photoinduced Transcription by Using Temporarily Mismatched Caged Oligonucleotides. *Angewandte Chemie International Edition* **2005,** *44*, 471-473.

30. Tang, X.; Richards, J. L.; Peritz, A. E.; Dmochowski, I. J., Photoregulation of DNA polymerase I (Klenow) with caged fluorescent oligodeoxynucleotides. *Bioorganic & Medicinal Chemistry Letters* **2005,** *15*, 5303-5306.

31. Hagen, V.; Bendig, J.; Frings, S.; Eckardt, T.; Helm, S.; Reuter, D.; Kaupp, U. B., Highly Efficient and Ultrafast Phototriggers for cAMP and cGMP by Using Long-Wavelength UV/Vis-Activation. *Angewandte Chemie International Edition* **2001,** *40*, 1046-1048.

32. Pinheiro, A. V.; Baptista, P.; Lima, J. C., Light activation of transcription: photocaging of nucleotides for control over RNA polymerization. *Nucleic Acids Research* **2008,** *36* (14), 1-7.

33. Deibel, Jr. M. R.; Coleman, M. S., Biochemical Properties of Purified Human Terminal Deoxynucleotidyltransferase. *The Journal of Biological Chemistry* **1980,** *255* (9), 4206-4212.

34. Harriman, W. D.; Wabl, M., A Video Technique for the Quantification of DNA in Gels Stained with Ethidium Bromide. *Analytical Biochemistry* **1995,** *228*, 336-342.

35. Boosalis, M. S.; Petruska, J., Goodman, M. F., DNA Polymerase Insertion Fidelity - Gel Assay for Site-Specific Kinetics. *The Journal of Biological Chemistry* **1987,** *262* (30), 14689-1469.

36. Boulé, J.-B.; Rougeon, F.; Papanicolaou, C., Terminal Deoxynucleotidyl Transferase Indiscriminately Incorporates Ribonucleotides and Deoxyribonucleotides. *The Journal of Biological Chemistry* **2001,** *276* (33), 31388–31393.
Pag37